PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/48, C07K 14/15, 16/10, G01N 33/569, 33/574, 33/577, C12Q 1/70, A61K 39/21

A1

(11) International Publication Number: WO 97/25431

(43) International Publication Date: 17 July 1997 (17.07.97)

(21) International Application Number: PCT/US97/00398

(22) International Filing Date: 10 January 1997 (10.01.97)

(30) Priority Data:
08/587,329 10 January 1996 (10.01.96) US

(71) Applicant: CORIXA CORPORATION [US/US]; Suite 464, 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: FRUDAKIS, Tony, N.; P.O. Box 99232, Magnolia Station, 3211 West McGraw, Seattle, WA 98232-0232 (US). SMITH, John, M.; 203 - 116th Place SE, Everett, WA 98208 (US).

(74) Agents: MAKI, David, J. et al.; Seed and Berry L.L.P., 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CANCER

(57) Abstract

Compositions and methods for the detection and therapy of cancer are disclosed. The compounds provided include human endogenous retroviral sequences that are preferentially expressed in tumor tissue, as well as polypeptides encoded by such nucleotide sequences. Vaccines and pharmaceutical compositions comprising such compounds are also provided and may be used, for example, for the prevention and treatment of cancer. The polypeptides may also be used for the production of antibodies, which are useful for diagnosing and monitoring the progression of cancer in a patient.

NUCLEOTIDE SEQUENCE OF THE REPRESENTATIVE
BREAST-TUMOR SPECIFIC cDNA B18Ag1

```
TTA GAG ACC CAA TTG GGA CCT AAT TGG GAC CCA AAT TTC TCA AGT GGA    48
Leu Glu Thr Gln Leu Gly Pro Asn Trp Asp Pro Asn Phe Ser Ser Gly
  1               5                  10                  15

GGG AGA ACT TTT GAC GAT TTC CAC CGG TAT CTC CTC GTG GGT ATT CAG    96
Gly Arg Thr Phe Asp Asp Phe His Arg Tyr Leu Leu Val Gly Ile Gln
             20                  25                  30

GGA GCT GCC CAG AAA CCT ATA AAC TTG TCT AAG GCG ATT GAA GTC GTC   144
Gly Ala Ala Gln Lys Pro Ile Asn Leu Ser Lys Ala Ile Glu Val Val
         35                  40                  45

CAG GGG CAT GAT GAG TCA CCA GGA GTG TTT TTA GAG CAC CTC CAG GAG   192
Gln Gly His Asp Glu Ser Pro Gly Val Phe Leu Glu His Leu Gln Glu
     50                  55                  60

GCT TAT CGG ATT TAC ACC CCT TTT GAC CTG GCA GCC CCC GAA AAT AGC   240
Ala Tyr Arg Ile Tyr Thr Pro Phe Asp Leu Ala Ala Pro Glu Asn Ser
 65                  70                  75                  80

CAT GCT CTT AAT TTG GCA TTT GTG GCT CAG GCA GCC CCA GAT AGT AAA   288
His Ala Leu Asn Leu Ala Phe Val Ala Gln Ala Ala Pro Asp Ser Lys
                 85                  90                  95

AGG AAA CTC CAA AAA CTA GAG GGA TTT TGC TGG AAT GAA TAC CAG TCA   336
Arg Lys Leu Gln Lys Leu Glu Gly Phe Cys Trp Asn Glu Tyr Gln Ser
            100                 105                 110

GCT TTT AGA GAT AGC CTA AAA GGT TTT                               363
Ala Phe Arg Asp Ser Leu Lys Gly Phe
        115                 120
```

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

Description

COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CANCER

Technical Field

The present invention relates generally to the detection and therapy of cancer. The invention is more specifically related to nucleotide sequences that are preferentially expressed in a tumor tissue and to polypeptides encoded by such nucleotide sequences. The invention is more particularly related to nucleotide sequences comprising at least a portion of a human endogenous retroviral sequence that is preferentially expressed in a tumor tissue, and to polypeptides encoded by such nucleotide sequences. The nucleotide sequences and polypeptides may be used in vaccines and pharmaceutical compositions for the prevention and treatment of cancer. The polypeptides may also be used for the production of compounds, such as antibodies, useful for diagnosing and monitoring the progression of cancer in a patient.

Background of the Invention

In recent years, considerable research has been directed to the identification of tumor markers, which may be useful for the diagnosis of particular cancers, for predicting the outcome of the disease or for developing a therapy in a patient-specific manner. Such research has generally focused on oncogenes, which are normal cellular genes whose expression has been altered (*e.g.*, by gene amplification, increased transcription, alteration of mRNA splicing or mutation within the coding region) such that otherwise normal cells assume neoplastic growth behavior. To date, however, the established markers have had a limited utility, and their use often leads to a result that is difficult to interpret.

Management of cancer currently relies on a combination of early diagnosis and aggressive treatment, which may include one or more of a variety of treatments such as surgery, radiotherapy, chemotherapy and hormone therapy. However, current diagnostic methods often fail to detect a cancer until the disease has progressed to a state that is difficult to treat, and existing treatments often have serious side effects. The high mortality observed among cancer patients indicates that improvements are needed in the diagnosis and treatment of the disease.

Accordingly, there is a need in the art for improved tumor markers, and methods for therapy and diagnosis of cancer. The present invention fulfills these needs and further provides other related advantages.

Summary of the Invention

Briefly stated, this invention provides compositions and methods for the diagnosis and therapy of cancer. In one aspect, isolated DNA molecules are provided, comprising: (a) a human endogenous retroviral sequence, wherein the retroviral sequence is preferentially expressed in a tumor tissue; (b) a variant of the human endogenous retroviral sequence that contains one or more nucleotide substitutions, deletions, insertions and/or modifications at no more than 20% (preferably no more than 5%) of the nucleotide positions, such that the antigenic and/or immunogenic properties of the polypeptide encoded by the human endogenous retroviral sequence are retained; or (c) a nucleotide sequence encoding an epitope of a polypeptide encoded by at least one of the above sequences. Isolated DNA and RNA molecules comprising a nucleotide sequence complementary to a DNA molecule as described above are also provided.

In another aspect, the present invention provides an isolated DNA molecule encoding an epitope of a polypeptide, the polypeptide being encoded by: (a) a nucleotide sequence transcribed from the sequence of SEQ ID NO:11; or (b) a variant of the nucleotide sequence that contains one or more nucleotide substitutions, deletions, insertions and/or modifications at not more than 20% of the nucleotide positions, such that the antigenic and/or immunogenic properties of the polypeptide encoded by the nucleotide sequence are retained. Isolated DNA and RNA molecules comprising a nucleotide sequence complementary to a DNA molecule as described above are also provided.

In related aspects, the present invention provides recombinant expression vectors comprising a DNA molecule as described above and host cells transformed or transfected with such expression vectors.

In further aspects, polypeptides, comprising an amino acid sequence encoded by a DNA molecule as described above, and monoclonal antibodies that bind to such polypeptides are provided.

In another aspect, methods are provided for determining the presence of a cancer in a patient. In one embodiment, the method comprises detecting, within a biological sample obtained from a patient, a polypeptide as described above. In another embodiment, the method comprises detecting, within a biological sample, an RNA molecule encoding a polypeptide as described above. In yet another embodiment, the method comprises (a) intradermally injecting a patient with a polypeptide as described above; and (b) detecting an immune response on the patient's skin and therefrom detecting the presence of a cancer in the patient.

In a related aspect, diagnostic kits useful in the determination of breast cancer are provided. The diagnostic kits generally comprise one or more monoclonal antibodies as described above, and a detection reagent. Within another related aspect, the diagnostic kit comprises a first polymerase chain reaction primer and a second polymerase chain reaction primer, the first and second primers each comprising at least about 10 contiguous nucleotides of an RNA molecule encoding a polypeptide as described above. Within yet another related aspect, the diagnostic kit comprises at least one oligonucleotide probe, the probe comprising at least about 15 contiguous nucleotides of a DNA molecule as described above. In another aspect, the present invention provides methods for monitoring the progression of a cancer in a patient. In one embodiment, the method comprises: (a) detecting an amount, in a biological sample, of a polypeptide as described above; (b) subsequently repeating step (a); and (c) comparing the amounts of polypeptide detected in steps (a) and (b), and therefrom monitoring the progression of cancer in the patient. In another embodiment, the method comprises (a) detecting an amount, within a biological sample, of an RNA molecule encoding a polypeptide as described above; (b) subsequently repeating step (a); and (c) comparing the amounts of RNA molecules detected in steps (a) and (b), and therefrom monitoring the progression of cancer in the patient.

In other aspects, pharmaceutical compositions, which comprise a polypeptide as described above and a physiologically acceptable carrier, and vaccines, which comprise a polypeptide as described above and an immune response enhancer are provided.

In related aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition or vaccine as described above.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

Brief Description of the Drawings

Figure 1 shows the differential display PCR products, separated by gel electrophoresis, obtained from cDNA prepared from normal breast tissue (lanes 1 and 2) and from cDNA prepared from breast tumor tissue from the same patient (lanes 3 and 4). The arrow indicates the band corresponding to B18Ag1.

Figure 2 is a northern blot comparing the level of B18Ag1 mRNA in breast tumor tissue (lane 1) with the level in normal breast tissue.

Figure 3 shows the level of B18Ag1 mRNA in breast tumor tissue compared to that in various normal and non-breast tumor tissues as determined by RNase protection assays.

Figure 4 is a genomic clone map showing the location of additional retroviral sequences (provided in SEQ ID NO:3 - SEQ ID NO:10) relative to B18Ag1.

Figures 5A and 5B show the sequencing strategy, genomic organization, and predicted open reading frame for the retroviral element containing B18Ag1.

Figure 6 shows the nucleotide sequence of the representative human endogenous retroviral element B18Ag1.

Detailed Description of the Invention

As noted above, the present invention is generally directed to compositions and methods for the diagnosis, monitoring and therapy of cancer. The compositions described herein include polypeptides, nucleic acid sequences and antibodies. Polypeptides of the present invention generally comprise at least a portion of a protein that is encoded by a human endogenous retroviral sequence, wherein the human endogenous retroviral sequence is expressed at substantially greater levels in a human tumor tissue than in normal tissue (*i.e.*, the level of RNA encoding the polypeptide is at least two fold higher, and preferably at least five fold higher, in a tumor tissue than in normal tissue). Such sequences are said to be "preferentially expressed" in a tumor tissue. Any cancer characterized by increased expression of a human endogenous retroviral sequence within a tumor may be detected and/or treated according to the present invention. Representative cancers include breast cancer, prostate cancer, leukemia, lymphoma and Kaposi's sarcoma. As used herein, the term "polypeptide" encompasses amino acid chains of any length, including full length proteins (and epitopes thereof) encoded by a human endogenous retroviral sequence.

Nucleic acid sequences of the subject invention generally comprise a DNA or RNA sequence that encodes a polypeptide as described above, or that is complementary to such a sequence. Antibodies are generally immune system proteins, or fragments thereof, that are capable of binding to a portion of a polypeptide as described above. Antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies.

Polypeptides within the scope of this invention include, but are not limited to, polypeptides (and epitopes thereof) encoded by the human endogenous retroviral sequences described herein. Such sequences include the sequence designated

B18Ag1 (SEQ ID NO:1) as well as other sequences such as those recited in SEQ ID NO:3-SEQ ID NO:10, found within the retroviral genome containing B18Ag1 (SEQ ID NO:11). B18Ag1 has homology to the P30 gene of the endogenous human retroviral element S71, as described in Werner et al., *Virology 174*:225-238 (1990). As used herein, the term "polypeptide" encompasses amino acid chains of any length, including full length proteins encoded by a human endogenous retroviral element. A polypeptide comprising an epitope of a human endogenous retroviral element may consist entirely of the epitope, or may contain additional sequences. The additional sequences may be derived from the native protein or may be heterologous, and such sequences may (but need not) possess immunogenic or antigenic properties.

An "epitope," as used herein is a portion of a polypeptide that is recognized (*i.e.*, specifically bound) by a B-cell and/or T-cell surface antigen receptor. Epitopes may generally be identified using well known techniques, such as those summarized in Paul, Fundamental Immunology, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides derived from the native polypeptide for the ability to react with antigen-specific antisera and/or T-cell lines or clones. An epitope of a polypeptide is a portion that reacts with such antisera and/or T-cells at a level that is similar to the reactivity of the full length polypeptide (e.g., in an ELISA and/or T-cell reactivity assay). Such screens may generally be performed using methods well known to those of ordinary skill in the art, such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. B-cell and T-cell epitopes may also be predicted via computer analysis. Polypeptides comprising an epitope of a polypeptide that is preferentially expressed in a tumor tissue (with or without additional amino acid sequence) are within the scope of the present invention.

The compositions and methods of the present invention also encompass variants of the above polypeptides and nucleic acid sequences encoding such polypeptides. A polypeptide "variant," as used herein, is a polypeptide that differs from the native polypeptide in substitutions and/or modifications such that the antigenic and/or immunogenic properties of the polypeptide are retained. Such variants may generally be identified by modifying one of the above polypeptide sequences and evaluating the reactivity of the modified polypeptide with antisera and/or T-cells as described above. Nucleic acid variants may contain one or more substitutions, deletions, insertions and/or modifications such that the antigenic and/or immunogenic properties of the encoded polypeptide are retained. One preferred variant of a human endogenous retroviral sequence, or an epitope thereof, is a variant that contains

nucleotide substitutions, deletions, insertions and/or modifications at no more than 20% of the nucleotide positions within the native polypeptide sequence.

Preferably, a variant contains conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenic or antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Human endogenous retroviral sequences that are expressed at substantially greater levels in a human tumor tissue than in normal tissue may be prepared using any of several techniques. For example, the human endogenous retroviral sequence designated B18Ag1 (Figure 6 and SEQ ID NO:1) may be cloned on the basis of its breast tumor specific expression, using differential display PCR. This technique compares the amplified products from poly A+ or total RNA template prepared from normal and breast tumor tissue. cDNA may be prepared by reverse transcription of RNA using a $(dT)_{12}AG$ primer. Following amplification using the primer CCTCAACCTC (SEQ ID NO:13), a band corresponding to an amplified product specific to the tumor RNA may be cut out from a silver stained gel and subcloned into a suitable vector (*e.g.*, the T-vector, Novagen, Madison, WI).

Alternatively, the B18Ag1 gene (or a portion thereof) may be amplified from human genomic DNA, or from breast tumor cDNA, via polymerase chain reaction. For this approach. B18Ag1 sequence-specific primers may be designed based on the sequence provided in SEQ ID NO:1, and may be purchased or synthesized. One suitable primer pair for amplification from breast tumor cDNA is (5'ATG GCT ATT TTC GGG GGC TGA CA) (SEQ ID NO:14) and (5'CCG GTA TCT CCT CGT GGG TAT T) (SEQ ID NO:15). An amplified portion of B18Ag1 may then be used to isolate

the full length gene from a human genomic DNA library or from a breast tumor cDNA library, using well known techniques such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (1989). Other sequences within the retroviral genome containing B18Ag1, such as those recited in SEQ ID NO:3 - SEQ ID NO:10, may be similarly prepared by screening human genomic libraries using B18Ag1-specific sequences as probes.

Other human endogenous retroviral sequences that are expressed at substantially greater levels in a human tumor tissue than in normal tissue may be prepared using methods known to those of ordinary skill in the art. For example, such sequences may be identified using low stringency hybridization, followed by PCR to identify conserved motifs. The level of expression in tumor tissue may generally be evaluated using the methods described herein, such as PCR and Northern blot analysis.

Recombinant polypeptides encoded by the DNA sequences described above may be readily prepared from the DNA sequences. For example, supernatants from suitable host/vector systems which secrete recombinant polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

In general, any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO.

Such techniques may also be used to prepare polypeptides comprising epitopes or variants of the native polypeptides. For example, variants of a native polypeptide may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis, and sections of the DNA sequence may be removed to permit preparation of truncated polypeptides. Portions and other variants having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing

amino acid chain. *See* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions.

In specific embodiments, polypeptides of the present invention encompass polypeptides encoded by a human endogenous retroviral sequence that is expressed at substantially greater levels in a human tumor tissue than in normal tissue (such as the sequence recited in SEQ ID NO:1), variants of such polypeptides that are encoded by DNA molecules containing one or more nucleotide substitutions, deletions, insertions and/or modifications at no more than 20% of the nucleotide positions, and epitopes of the above polypeptides. Polypeptides within the scope of the present invention also include polypeptides (and epitopes thereof) encoded by DNA sequences that hybridize to the above sequences under stringent conditions, wherein the DNA sequences are at least 80% identical in overall sequence to the sequence recited in SEQ ID NO:1, and wherein RNA corresponding to said nucleotide sequence is expressed at a greater level in human tumor tissue than in the corresponding normal tissue. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing overnight at 65°C in 6X SSC, 0.2% SDS; followed by washing twice at 65° C for 30 minutes each with 1X SSC, 0.1% SDS, and then washing twice at 65°C for 30-60 minutes each with 0.1X SSC, 0.1% SDS. DNA molecules according to the present invention include molecules that encode any of the above polypeptides.

In another aspect of the present invention, antibodies are provided. Such antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for the antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J.*

Immunol. 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Antibodies may be used, for example, in methods for detecting a cancer (such as breast cancer, prostate cancer, leukemia, lymphoma or Kaposi's sarcoma) in a patient. Such methods involve using one or more antibodies to detect the presence or absence of a polypeptide as described herein in a suitable biological sample. As used herein, suitable biological samples include tumor or normal tissue biopsy, mastectomy, blood, lymph node, serum and urine samples or other tissue, homogenate or extract thereof, obtained from a patient. It will be evident to those of ordinary skill in the art that, following detection of a polypeptide within a non-biopsy sample, additional tumor markers may be employed to identify the particular type of cancer.

There are a variety of assay formats known to those of ordinary skill in the art for using an antibody to detect polypeptide markers in a sample. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. For example, the assay may be performed in a Western blot format, wherein a protein preparation from the biological sample is submitted to gel electrophoresis, transferred to a suitable membrane and allowed to react with antibody. The presence of

antibody on the membrane may then be detected using a suitable detection reagent, as described below.

In another embodiment, the assay involves the use of an antibody immobilized on a solid support to bind to the polypeptide and remove it from the remainder of the sample. The bound polypeptide may then be detected using a second antibody that binds to the binding partner/polypeptide complex and contains a reporter group. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized antibody after incubation of the antibody with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the antibody is indicative of the reactivity of the sample with the immobilized antibody, and as a result is indicative of the concentration of polypeptide in the sample.

The solid support may be any material known to those of ordinary skill in the art to which the antibody may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose filter or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The antibody may be immobilized on the solid support using a variety of techniques known to those in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the antibody, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of antibody ranging from about 10 ng to about 1 μg, and preferably about 100-200 ng, is sufficient to immobilize an adequate amount of polypeptide.

Covalent attachment of antibody to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the antibody. For example, the antibody may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde

group on the support with an amine and an active hydrogen on the binding partner (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook (1991) at A12-A13).

In certain embodiments for detection of polypeptide in a sample, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the biological sample, such that the polypeptide within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a second antibody (containing a reporter group) capable of binding to a different site on the polypeptide is added. The amount of second antibody that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is that period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with breast cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of antibody to reporter group may be achieved using standard methods known to those of ordinary skill in the art.

The second antibody is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound second antibody is then removed

and bound second antibody is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value may be considered positive for a cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, p. 106-7 (Little Brown and Co., 1985). Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the antibody is immobilized on a membrane, such as nitrocellulose. In the flow-through test, the polypeptide within the sample binds to the immobilized antibody as the sample passes through the membrane. A second, labeled antibody then binds to the antibody-polypeptide complex as a solution containing the second antibody flows through the membrane. The detection of bound second antibody may then be performed as described above. In the strip test format, one end of the membrane to which antibody is bound is immersed in a solution containing the sample.

The sample migrates along the membrane through a region containing second antibody and to the area of immobilized antibody. Concentration of second antibody at the area of immobilized antibody indicates the presence of breast cancer. Typically, the concentration of second antibody at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of antibody immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1μg, and more preferably from about 50 ng to about 1μg. Such tests can typically be performed with a very small amount of biological sample.

The presence or absence of a cancer in a patient may also be determined by evaluating the level of mRNA encoding a polypeptide of the present invention within the biological sample (*e.g.*, a biopsy, mastectomy and/or blood sample from a patient) relative to a predetermined cut-off value. Such an evaluation may be achieved using any of a variety of methods known to those of ordinary skill in the art such as, for example, *in situ* hybridization and amplification by polymerase chain reaction. For example, polymerase chain reaction may be used to amplify sequences from cDNA prepared from RNA that is isolated from one of the above biological samples. Sequence-specific primers for use in such amplification may be designed based on a cDNA or genomic sequence, such as a sequence provided in SEQ ID NO:1 or SEQ ID NO:3 - SEQ ID NO:10, and may be purchased or synthesized. In the case of B18Ag1, as noted herein, one suitable primer pair is (5'ATG GCT ATT TTC GGG GGC TGA CA) (SEQ ID NO:14) and (5'CCG GTA TCT CCT CGT GGG TAT T) (SEQ ID NO:15). The PCR reaction products may then be separated and visualized using gel electrophoresis, according to methods well known to those of ordinary skill in the art. Amplification is typically performed on samples obtained from matched pairs of tissue (tumor and non-tumor tissue from the same individual) or from unmatched pairs of tissue (tumor and non-tumor tissue from different individuals). The amplification reaction is preferably performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the tumor sample as compared to the same dilution of the non-tumor sample is considered positive.

Conventional RT-PCR protocols using agarose and ethidium bromide staining, while important in defining gene specificity do not lend themselves to diagnostic kit development because of the time and effort required in making them

quantitative (*i.e.*, construction of saturation and/or titration curves), and their sample throughput. This problem is overcome by the development of procedures such as real time RT-PCR which allows for assays to be performed in single tubes, and in turn can be modified for use in 96 well plate formats. Instrumentation to perform such methodologies are available from ABI/Perkin Elmer. Alternatively, other high throughput assays using labelled probes (*e.g.*, digoxygenin) in combination with labelled (*e.g.*, enzyme fluorescent, radioactive) antibodies to such probes can also be used in the development of 96 well plate assays.

In yet another method for determining the presence or absence of a cancer in a patient, one or more of the polypeptides described above may be used in a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to a test antigen (*i.e.*, an immunogenic portion of a polypeptide employed, or a variant thereof). The response may measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of a cancer. As noted above, additional tumor markers may be employed, using methods known to those of ordinary skill in the art, to identify the type of cancer present.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing at least one polypeptide and a physiologically acceptable carrier, such as water, saline, alcohol, or a buffer. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 1 μg to 100 μg, preferably from about 10 μg to 50 μg in a volume of 0.1 mL. Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80™.

In other aspects of the present invention, the progression and/or response to treatment of a cancer may be monitored by performing any of the above assays over a period of time, and evaluating the change in the level of the response (*i.e.*, the amount of polypeptide or mRNA detected or, in the case of a skin test, the extent of the immune response detected). For example, the assays may be performed every 1-2 months for a

period of 1-2 years. In general, a cancer is progressing in those patients in whom the level of the response increases over time. In contrast, a cancer is not progressing when the signal detected either remains constant or decreases with time.

In further aspects of the present invention, the compounds described herein may be used for the immunotherapy of a cancer. In these aspects, the compounds (which may be polypeptides, antibodies or nucleic acid molecules) are preferably incorporated into pharmaceutical compositions or vaccines. Pharmaceutical compositions comprise one or more such compounds and a physiologically acceptable carrier. Vaccines may comprise one or more polypeptides and an immune response enhancer, such as an adjuvant or a liposome (into which the compound is incorporated). Pharmaceutical compositions and vaccines may additionally contain a delivery system, such as biodegradable microspheres which are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109. Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, including one or more separate polypeptides.

Alternatively, a vaccine may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. In such vaccines, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (*e.g.*, vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., *Science 259*:1745-1749 (1993) and reviewed by Cohen, *Science 259*:1691-1692 (1993). The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier,

such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention.

Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, *Bordella pertussis* or *Mycobacterium tuberculosis*-derived proteins. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI), Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ), alum, biodegradable microspheres, monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

The above pharmaceutical compositions and vaccines may be used, for example, for the therapy of cancer in a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient may or may not be afflicted with a cancer. Accordingly, the above pharmaceutical compositions and vaccines may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. To prevent the development of a cancer, a pharmaceutical composition or vaccine comprising one or more polypeptides as described herein (or naked, plasmid or viral vector DNA encoding such a polypeptide) may be administered to a patient. For treating a patient with a cancer, the pharmaceutical composition or vaccine may comprise one or more polypeptides, antibodies or nucleic acid molecules complementary to DNA encoding a polypeptide as described herein (*e.g.*, antisense RNA or antisense deoxyribonucleotide oligonucleotides).

For example, tumor cells that express a polypeptide as described herein may be preferentially killed by administering to a patient a conjugate in which a cytotoxic agent or "prodrug" is linked to antisense RNA, an antisense deoxyribonucleotide oligonucleotide or an antibody that binds to such a polypeptide. As used herein, the term "prodrug" refers to a group that is not itself toxic to the cell, but that can be rendered toxic after the conjugate is directed to the target cell by the addition of a second activating compound, such as an enzyme that can convert the prodrug into an active drug. Any suitable cytotoxic agent (including radionuclides) or prodrug known to those of ordinary skill in the art may be employed in such methods. Suitable prodrugs include boron, doxifluridine, or the prodrug precursor of palytoxin.

Routes and frequency of administration, as well as dosage, will vary from individual to individual. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Between 1 and 10 doses may be administered for a 52 week period. Preferably, 6 doses are administered, at intervals of one month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response. Such a response can be monitored by measuring the level of anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. A suitable dose should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients.. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 100 μg to about 5 mg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1

Preparation of B18Ag1 cDNA and Genomic Clones Using Differential Display RT-PCR

This Example illustrates the preparation of cDNA and genomic DNA molecules encoding B18Ag1 using a differential display screen.

Tissue samples were prepared from breast tumor and normal tissue of a patient with breast cancer that was confirmed by pathology after removal from the patient. Normal RNA and tumor RNA was extracted from the samples and mRNA was isolated and converted into cDNA using a $(dT)_{12}AG$ anchored 3' primer. Differential display PCR was then executed using a randomly chosen primer (CTTCAACCTC) (SEQ ID NO:16). Amplification conditions were standard buffer containing 1.5 mM $MgCl_2$, 20 pmol of primer, 500 pmol dNTP, and 1 unit of *Taq* DNA polymerase (Perkin-Elmer, Branchburg, NJ). Forty cycles of amplification were performed using 94°C denaturation for 30 seconds, 42°C annealing for 1 minute, and 72°C extension for 30 seconds. An RNA fingerprint containing 76 amplified products was obtained. Although the RNA fingerprint of breast tumor tissue was over 98% identical to that of the normal breast tissue, a band was repeatedly observed to be specific to the RNA fingerprint pattern of the tumor. This band was cut out of a silver stained gel and subcloned into the T-vector (Novagen, Madison, WI) and sequenced.

The sequence of the cDNA, referred to as B18Ag1, is provided in SEQ ID NO:1. A database search of GENBANK and EMBL revealed that the B18Ag1 fragment initially cloned is 77% identical to the endogenous human retroviral element S71, which is a truncated retroviral element homologous to the Simian Sarcoma Virus (SSV). S71 contains a complete *gag* gene, a portion of the *pol* gene and an LTR-like structure at the 3' terminus (*see* Werner et al., *Virology 174*:225-238 (1990)). B18Ag1 is also 64% identical to SSV in the region corresponding to the P30 (gag) locus. B18Ag1 contains three separate and incomplete reading frames covering a region which shares considerable homology to a wide variety of gag proteins of retroviruses which infect mammals. In addition, the homology to S71 is not just within the *gag* gene, but spans several kb of sequence including an LTR.

B18Ag1-specific PCR primers were synthesized using computer analysis guidelines. RT-PCR amplification (94°C, 30 seconds; 60°C → 42°C, 30 seconds; 72°C, 30 seconds, for 40 cycles) confirmed that B18Ag1 represents an actual mRNA sequence present at relatively high levels in the patient's breast tumor tissue.

The primers used in amplification were B18Ag1-1 (CTG CCT GAG CCA CAA ATG) (SEQ ID NO:17) and B18Ag1-4 (CCG GAG GAG GAA GCT AGA GGA ATA) (SEQ ID NO:18) at a 3.5 mM magnesium concentration and a pH of 8.5, and B18Ag1-2 (ATG GCT ATT TTC GGG GGC TGA CA) (SEQ ID NO:14) and B18Ag1-3 (CCG GTA TCT CCT CGT GGG TATT) (SEQ ID NO:15) at 2 mM magnesium at pH 9.5. The same experiments showed exceedingly low to nonexistent levels of expression in this patient's normal breast tissue (*see* Figure 1). RT-PCR experiments were then used to show that B18Agl mRNA is present in nine other breast tumor samples (from Brazilian and American patients) but absent in, or at exceedingly low levels in, the normal breast tissue corresponding to each cancer patient. RT-PCR analysis has also shown that the B18Agl transcript is not present in various normal tissues (including lymph node, myocardium and liver) and present at relatively low levels in PBMC and lung tissue. The presence of B18Agl mRNA in breast tumor samples, and its absence from normal breast tissue, has been confirmed by Northern blot analysis, as shown in Figure 2.

The differential expression of B18Ag1 in breast tumor tissue was also confirmed by RNase protection assays. Figure 3 shows the level of B18Ag1 mRNA in various tissue types as determined in four different RNase protection assays. Lanes 1-12 represent various normal breast tissue samples, lanes 13-25 represent various breast tumor samples; lanes 26-27 represent normal prostate samples; lanes 28-29 represent prostate tumor samples; lanes 30-32 represent colon tumor samples; lane 33 represents normal aorta; lane 34 represents normal small intestine; lane 35 represents normal skin, lane 36 represents normal lymph node; lane 37 represents normal ovary; lane 38 represents normal liver; lane 39 represents normal skeletal muscle; lane 40 represents a first normal stomach sample, lane 41 represents a second normal stomach sample; lane 42 represents a normal lung; lane 43 represents normal kidney; and lane 44 represents normal pancreas. Interexperimental comparison was facilitated by including a positive control RNA of known B-actin message abundance in each assay and normalizing the results of the different assays with respect to this positive control.

RT-PCR and Southern blot analysis has shown the B18Ag1 locus to be present in human genomic DNA as a single copy endogenous retroviral element. A genomic clone of approximately 12-18 kb was isolated using the initial B18Ag1 sequence as a probe. Four additional subclones were also isolated by XbaI digestion. Additional retroviral sequences obtained from these clones (located as shown in Figure 4) are shown as SEQ ID NO:3 - SEQ ID NO:10, where SEQ ID NO:3 shows the location of the sequence labeled 10 in Figure 4, SEQ ID NO:4 shows the location of the sequence labeled 11-29, SEQ ID NO:5 shows the location of the sequence labeled 3,

SEQ ID NO:6 shows the location of the sequence labeled 6, SEQ ID NO:7 shows the location of the sequence labeled 12, SEQ ID NO:8 shows the location of the sequence labeled 13, SEQ ID NO:9 shows the location of the sequence labeled 14 and SEQ ID NO:10 shows the location of the sequence labeled 11-22.

Subsequent studies demonstrated that the 12-18 kb genomic clone contains a retroviral element of about 7.75 kb, as shown in Figures 5A and 5B. The sequence of this retroviral element is shown in SEQ ID NO:11. The numbered line at the top of Figure 5A represents the sense strand sequence of the retroviral genomic clone. The box below this line shows the position of selected restriction sites. The arrows depict the different overlapping clones used to sequence the retroviral element. The direction of the arrow shows whether the single-pass subclone sequence corresponded to the sense or anti-sense strand. Figure 5B is a schematic diagram of the retroviral element containing B18Ag1 depicting the organization of viral genes within the element. The open boxes correspond to predicted reading frames, starting with a methionine, found throughout the element. Each of the six likely reading frames is shown, as indicated to the left of the boxes, with frames 1-3 corresponding to those found on the sense strand.

Using the cDNA of SEQ ID NO:1 as a probe, a longer cDNA was obtained (SEQ ID NO:12) which contains minor nucleotide differences (less than 1%) compared to the genomic sequence shown in SEQ ID NO:11.

Example 2

Preparation of B18Ag1 DNA from Human Genomic DNA

This example illustrates the preparation of B18Ag1 DNA by amplification from human genomic DNA.

B18Ag1 DNA may be prepared from 250 ng human genomic DNA using 20 pmol of B18Ag1 specific primers, 500 pmol dNTPS and 1 unit of *Taq* DNA polymerase (Perkin Elmer, Branchburg, NJ) using the following amplification parameters: 94°C denaturing for 30 seconds, 30 second 60°C to 42°C touchdown annealing in 2°C increments every two cycles and 72°C extension for 30 seconds. The last increment (a 42°C annealing temperature) should cycle 25 times. Primers (B18Ag1-1, B18Ag1-2, B18Ag1-3 and B18Ag1-4) were selected using computer analysis. Primers synthesized were. Primer pairs that may be used are 1+3, 1+4, 2+3, and 2+4.

Following gel electrophoresis, the band corresponding to B18Ag1 DNA may be excised and cloned into a suitable vector.

Example 3
Preparation of B18Ag1 DNA from Breast Tumor cDNA

This example illustrates the preparation of B18Ag1 DNA by amplification from human breast tumor cDNA.

First strand cDNA is synthesized from RNA prepared from human breast tumor tissue in a reaction mixture containing 500 ng poly A+ RNA, 200 pmol of the primer (T)12AG (*i.e.*, TTT TTT TTT TTT AG) (SEQ ID NO:19), 1X first strand reverse transcriptase buffer, 6.7 mM DTT, 500 mmol dNTPs, and 1 unit AMV or MMLV reverse transcriptase (from any supplier, such as Gibco-BRL (Grand Island, NY)) in a final volume of 30 µl. After first strand synthesis, the cDNA is diluted approximately 25 fold and 1 µl is used for amplification as described in Example 2. While some primer pairs can result in a heterogeneous population of transcripts, the primers B18Ag1-2 (5'ATG GCT ATT TTC GGG GGC TGA CA) (SEQ ID NO:14) and B18Ag1-3 (5'CCG GTA TCT CCT CGT GGG TAT T) (SEQ ID NO:15) yield a single 151 bp amplification product.

Example 4
Identification of B-cell and T-cell Epitopes of B18Ag1

This Example illustrates the identification of B18Ag1 epitopes.

The B18Ag1 sequence can be screened using a variety of computer algorithms. To determine B-cell epitopes, the sequence can be screened for hydrophobicity and hydrophilicity values using the method of Hopp, *Prog. Clin. Biol. Res. 172B*:367-77 (1985) or, alternatively, Cease et al., *164 J. Exp. Med.* 1779-84 (1986) or Spouge et al., *J. Immunol. 138*:204-12 (1987). Additional Class II MHC (antibody or B-cell) epitopes can be predicted using programs such as AMPHI (*e.g.*, Margalit et al., *J. Immunol. 138*:2213 (1987)) or the methods of Rothbard and Taylor (*e.g.*, *EMBO J.* 7:93 (1988)).

Once peptides (15-20 amino acids long) are identified using these techniques, individual peptides can be synthesized using automated peptide synthesis equipment (available from manufacturers such as Applied BioSystems, Inc., Foster City, CA) and techniques such as Merrifield synthesis. Following synthesis, the peptides can used to screen sera harvested from either normal or breast cancer patients to determine whether patients with breast cancer possess antibodies reactive with the peptides. Presence of such antibodies in breast cancer patient would confirm the

immunogenicity of the specific B-cell epitope in question. The peptides can also be tested for their ability to generate a serologic or humoral immune in animals (mice, rats, rabbits, chimps etc.) following immunization *in vivo*. Generation of a peptide-specific antiserum following such immunization further confirms the immunogenicity of the specific B-cell epitope in question.

To identify T-cell epitopes, the B18Ag1 sequence can be screened using different computer algorithms which are useful in identifying 8-10 amino acid motifs within the B18Ag1 sequence which are capable of binding to HLA Class I MHC molecules. (*see, e.g.*, Rammensee et al., *Immunogenetics 41*:178-228 (1995)). Following synthesis such peptides can be tested for their ability to bind to class I MHC using standard binding assays (*e.g.*, Sette et al., *J. Immunol. 153*:5586-92 (1994)) and more importantly can be tested for their ability to generate antigen reactive cytotoxic T-cells following *in vitro* stimulation of patient or normal peripheral mononuclear cells using, for example, the methods of Bakker et al., *Cancer Res. 55*:5330-34 (1995); Visseren et al., *J. Immunol. 154*:3991-98 (1995); Kawakami et al., *J. Immunol. 154*:3961-68 (1995); and Kast et al., *J. Immunol. 152*:3904-12 (1994). Successful *in vitro* generation of T-cells capable of killing autologous (bearing the same class I MHC molecules) tumor cells following *in vitro* peptide stimulation further confirms the immunogenicity of the B18Ag1 antigen. Furthermore, such peptides may be used to generate murine peptide and B18Ag1 reactive cytotoxic T-cells following *in vivo* immunization in mice rendered transgenic for expression of a particular human MHC Class I haplotype (Vitiello et al., *J. Exp. Med. 173*:1007-15 (1991).

A representative a list of predicted B18Ag1 B-cell and T-cell epitopes, broken down according to predicted HLA Class I MHC binding antigen, is shown below:

<u>Predicted Th Motifs (B-cell epitopes)</u>

SSGGRTFDDFHRYLLVGI (SEQ ID NO:20)
QGAAQKPINLSKXIEVVQGHDE (SEQ ID NO:21)
SPGVFLEHLQEAYRIYTPFDLSA (SEQ ID NO:22)

<u>Predicted HLA A2.1 Motifs (T-cell epitopes)</u>

YLLVGIQGA (SEQ ID NO:23)
GAAQKPINL (SEQ ID NO:24)
NLSKXIEVV (SEQ ID NO:25)
EVVQGHDES (SEQ ID NO:26)
HLQEAYRIY (SEQ ID NO:27)

NLAFVAQAA (SEQ ID NO:28)
FVAQAAPDS (SEQ ID NO:29)

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Corixa Corporation

(ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CANCER

(iii) NUMBER OF SEQUENCES: 29

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center. 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7092

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0. Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 10-JAN-1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Maki. David J.
(B) REGISTRATION NUMBER: 31.392
(C) REFERENCE/DOCKET NUMBER: 210121.418PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (206) 622-4900
(B) TELEFAX: (206) 682-6031

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TTGANTGTCA AAAACCTTNT AGGCTATCTC TAAAAGCTGA CTGGTATTCA TTCCAGCAAA     60
ATCCCTCTAG TTTTTGGAGT TTCCTTTTAC TATCTGGGGC TGCCTGAGCC ACAAATGCCA    120
AATTAAGAGC ATGGCTATTT TCGGGGGCTG ACAGGTCAAA AGGGGTGTAA ATCCGATAAG    180
CCTCCTGGAG GTGCTCTAAA AACACTCCTG GTGACTCATC ATGCCCCTGG ACGACTTCAA    240
TCGNCTTAGA CAAGTTTATA GGTTTCTGGG CAGTCCCTGA ATACCCACGA GGAGATACCG    300
GTGGAAATCG TCAAAAGTTC TCCCTCCACT TGAGAAATTT GGGTCCCAAT TAGGTCCCAA    360
TTGGGTCTCT AATCACTATT CCTCTAGCTT CCTCCTCCGG NCTATTGGTT GATGT         415
```

(2) INFORMATION FOR SEQ ID NO:2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Trp Asp Pro Asn Phe Ser Ser Gly Gly Arg Thr Phe Asp Asp Phe His
1               5                   10                  15

Arg Tyr Leu Leu Val Gly Ile Gln Gly Ala Ala Gln Lys Pro Ile Asn
            20                  25                  30

Leu Ser Lys Xaa Ile Glu Val Val Gln Gly His Asp Glu Ser Pro Gly
        35                  40                  45

Val Phe Leu Glu His Leu Gln Glu Ala Tyr Arg Ile Tyr Thr Pro Phe
    50                  55                  60

Asp Lys Ser Ala Pro Glu Asn Ser His Ala Leu Asn Leu Ala Phe Val
65                  70                  75                  80

Ala Gln Ala Ala Pro Asp Ser Lys Arg Lys Leu Gln Lys Leu Glu Gly
                85                  90                  95
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1180 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
NCNNNNNTTA TGATTACGCC AAGCGNGCAA TTAACCCTCA CTAAAGGGAA CAAAAGCTGG   60
AGCTCCACCG CGGTGGCGGC CGCTAGAATC TTCATACCCC GAACTCTTGG GAAAACTTTA  120
ATCAGTCACC TACAGTCTAC CACCCATTTA GGAGGAGCAA AGCTACCTCA GCTCCTCCGG  180
AGCCGTTTTA AGATCCCCCA TCTTCAAAGC CTAACAGATC AAGCAGCTCT CCGGTGCACA  240
ACCTGCGCCC AGGTAAATGC CAAAAAAGGT CCTAAACCCA GCCCAGGCCA CCGTCTCCAA  300
GAAAACTCAC CAGGAGAAAA GTGGGAAATT GACTTTACAG AAGTAAAACC ACACCGGGCT  360
GGGTACAAAT ACCTTCTAGT ACTGGTAGAC ACCTTCTCTG GATGGACTGA AGCATTTGCT  420
ACCAAAAACG AAACTGTCAA TATGGTAGTT AAGTTTTTAC TCAATGAAAT CATCCCTCGA  480
CGTGGGCTGC CTGTTGCCAT AGGGTCTGAT AATGGAACGG CCTTCGCCTT GTCTATAGTT  540
TAATCAGTCA GTAAGGCGTT AAACATTCAA TGGAAGCTCC ATTGTGCCTA TCGACCCAGA  600
GCTCTGGGAA GTAGAACGCA TGAACTGCAC CCTAAAAAAA CACTCTTACA AAATTAATCT  660
TAAAAACCGG TGTTAATTGT GTTAGTCTCC TTCCCTTAGC CCTACTTAGA GTTAAGGTGC  720
ACCCCTTACT GGGCTGGGTT CTTTACCTTT TGAAATCATN TTTNGGAAGG GGCTGCCTAT  780
CTTTNCTTAA CTAAAAAANG CCCATTTGGC AAAAATTTCN CAACTAATTT NTACGTNCCT  840
ACGTCTCCCC AACAGGTANA AAAATCTNCT GCCCTTTTCA AGGAACCATC CCATCCATTC  900
```

CTNAACAAAA GGCCTGCCNT TCTTCCCCCA GTTAACTNTT TTTTNTTAAA ATTCCCAAAA 960

AANGAACCNC CTGCTGGAAA AACNCCCCCC TCCAANCCCC GGCCNAAGNG GAAGGTTCCC 1020

TTGAATCCCN CCCCCNCNAA NGGCCCGGAA CCNTTAAANT NGTTCCNGGG GGTNNGGCCT 1080

AAAAGNCCNA TTTGGTAAAC CTANAAATTT TTTCTTTTNT AAAAACCACN NTTTNNTTTT 1140

TCTTAAACAA AACCCTNTTT NTAGNANCNT ATTTCCCNCC 1180

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1163 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TNCTTTGATA CCCNAGCGTT CAATTAACCC TCACTAAAGG GAACAAAAGC TGGAGCTCCA 60

CCGCGGTGGC GGCCGCTCTA GAGCTGCGCC TGGATCCCGC CACAGTGAGG AGACCTGAAG 120

ACCAGAGAAA ACACAGCAAG TAGGCCCTTT AAACTACTCA CCTGTGTTGT CTTCTAATTT 180

ATTCTGTTTT ATTTTGTTTC CATCATTTTA AGGGGTTAAA ATCATCTTGT TCAGACCTCA 240

GCATATAAAA TGACCCATCT GTAGACCTCA GGCTCCAACC ATACCCCAAG AGTTGTCTGG 300

TTTTGTTTAA ATTACTGCCA GGTTTCAGCT GCAGATATCC CTGGAAGGAA TATTCCAGAT 360

TCCCTGAGTA GTTTCCAGGT TAAAATCCTA TAGGCTTCTT CTGTTTTGAG GAAGAGTTCC 420

TGTCAGAGAA AAACATGATT TTGGATTTTT AACTTTAATG CTTGTGAAAC GCTATAAAAA 480

AAATTTTCTA CCCCTAGCTT TAAAGTACTG TTAGTGAGAA ATTAAAATTC CTTCAGGAGG 540

ATTAAACTGC CATTTCAGTT ACCCTAATTC CAAATGTTTT GGTGGTTAGA ATCTTCTTTA 600

ATGTTCTTGA AGAAGTGTTT TATATTTTCC CATCNAGATA AATTCTCTCN CNCCTTNNTT 660

TTNTNTCTNN TTTTTTAAAA CGGANTCTTG CTCCGTTGTC CANGCTGGGA ATTTTNTTTT 720

GGCCAATCTC CGCTNCCTTG CAANAATNCT GCNTCCCAAA ATTACCNCCT TTTTCCCACC 780

TCCACCCCNN GGAATTACCT GGAATTANAG GCCCCCNCCC CCCCCCCGGC TAATTTGTTT 840

TTGTTTTTAG TAAAAAACGG GTTTCCTGTT TTAGTTAGGA TGGCCCANNT CTGACCCCNT 900

NATCNTCCCC CTCNGCCCTC NAATNTTNGG NNTANGGCTT ACCCCCCCCN GNNGTTTTTC 960

CTCCATTNAA ATTTTCTNTG GANTCTTGAA TNNCGGGTTT TCCCTTTTAA ACCNATTTTT 1020

TTTTTNNNNC CCCCANTTTT NCCTCCCCCN TNTNTAANGG GGGTTTCCCA ANCCGGGTCC 1080

NCCCCCANGT CCCCAATTTT TCTCCCCCCC CCTCTTTTTT CTTTNCCCCA AAANTCCTAT 1140

CTTTTCCTNN AAATATCNAN TNT 1163

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1122 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

NNGGTCCNNC TCAAAGTCAN TATAGGGCGA ATTGGGTACC GGGCCCCCCC TCGAGGTCGA 60

CGGTATCGAT AAGCTTGATA TCGAATTCCT GCAGCCCGGG GGATCCACTA GTTCTAGACC 120

AAGAAATGGA GGATTTTAGA GTGACTGATG ATTTCTCTAT CATCTGCAGT TAGTAAACAT 180

TCTCCACAGT TTATGCAAAA AGTAACAAAA CCACTGCAGA TGACAAACAC TAGGTAACAC 240

ACATACTATC TCCCAAATAC CTACCCACAA GCTCAACAAT TTTAAACTGT TAGGATCACT 300

GGCTCTAATC ACCATGACAT GAGGTCACCA CCAAACCATC AAGCGCTAAA CAGACAGAAT 360

GTTTCCACTC CTGATCCACT GTGTGGGAAG AAGCACCGAA CTTACCCACT GGGGGGCCTG 420

CNTCANAANA AAAGCCCATG CCCCCGGGTN TNCCTTTNAA CCGGAACGAA TNAACCCACC 480

ATCCCCACAN CTCCTCTGTT CNTGGGCCCT GCATCTTGTG GCCTCNTNTN CTTTNGGGGA 540

NACNTGGGGA AGGTACCCCA TTTCNTTGAC CCCNCNANAA AACCCCNGTG GCCCTTTGCC 600

CTGATTCNCN TGGGCCTTTT CTCTTTTCCC TTTTGGGTTG TTTAAATTCC CAATGTCCCC 660

NGAACCCTCT CCNTNCTGCC CAAAACCTAC CTAAATTNCT CNCTANGNNT TTTCTTGGTG 720

TTNCTTTTCA AAGGTNACCT TNCCTGTTCA NNCCCNACNA AAATTTNTTC CNTATNNTGG 780

NCCCNNAAAA ANNNATCNNC CCNAATTGCC CGAATTGGTT NGGTTTTTCC TNCTGGGGGA 840

AACCCTTTAA ATTTCCCCCT TGGCCGGCCC CCCTTTTTTC CCCCCTTTNG AAGGCAGGNR 900

GGTTCTTCCC GAACTTCCAA TTNCAACAGC CNTGCCCATT GNTGAAACCC TTTTCCTAAA 960

ATTAAAAAAT ANCCGGTTNN GGNNGGCCTC TTTCCCCTCC NGGNGGGNNG NGAAANTCCT 1020

TACCCCNAAA AAGGTTGCTT AGCCCCCNGT CCCCACTCCC CCNGGAAAAA TNAACCTTTT 1080

CNAAAAAAGG AATATAANTT TNCCACTCCT TNGTTCTCTT CC 1122

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1091 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

NCNNNCCNTT TGTNAAAGAC CGNCAGTGAG CGCGCGTAAT ACGACTCACT ATAGGGCGAA 60

TTGGGTACCG GGCCCCCCCT CGAGGTCGAC GGTATCGATA AGCTTGATAT CGAATTCCTG 120

CAGCCCGGGG GATCCACTAG TTCTAGAGCT CGCGGCCGCG AGCTCTAATA CGACTCACTA 180

TAGGGCGTCG ACTCGATCTC AGCTCACTGC AATCTCTGCC CCCGGGGTCA TGCGATTCTC 240

CTGCCTCAGC CTTCCAAGTA GCTGGGATTA CAGGCGTGCA ACACCACACC CGGCTAATTT 300

TGTATTTTTA ATAGAGATGG GGTTTTCCCT TGTTGGCCAN NATGGTCTCN AACCCCTGAC 360

CTCNNGTGAT CCCCCCNCCC NNGANCTCNN ACTGCTGGGG ATNNCCGNNN NNNNCCTCCC 420

NNCNCNNNNN NNCNCNNTCC NTNNTCCTTN CTCNNNNNNN NCNNTCNNTC CNNCTTCTCN 480

CCNNNTNTTN TCNNCNNCCN NCNNNCCNCN TNCCCNCNNN TTCNCNTNCN NTNTCCNNCN 540

NNNTCNNCNN NCNNNNCNTN NCCNNTACNT CNTNNNCNNN TCCNTCTNTN NCCTCNNCNN 600

TCNCTNCNCN TTNTCTCCTC NNTNNNNNNC TCCNNNNNTC TCNTCNCNNC NTNCCTCNNT 660

NNCCNCNCCC CNCCTCNCNN CCTNNTTTNN NCNNCNNNTC CNTNCCNTTC NNNTCCNNTN 720

NCNNCNTCNC NNNCNTTNTT CCCNCCNNTT CCTTNCNCNT NNNNTNTCNN NCNCNTCNNT 780

CNTTTNCTCC TNNNTCCCNN CTCNNTTCNC CCNNNTCCNC CCCCCNCCTN TCTCTCNCCC 840

NNNTNNNTNT NNNNCNTCCN CTNTCNCNTT CNTCNNTNCN TTNCTNTCNN CNNCNNTNCN 900

CTNCCNTNTN TCTNNNTCNC NTCNCNTNTC NCCNTCCNTT NCTNTCTCCT NTNTCCTTCC 960

CCTCNCCTNC TCNTTCNCCN CCCNNTNTNT NTNNCNCCNN TNCTNNNCNN CCNTCNTTTC 1020

NTCTCTNCTN NNNNTNNCCT CNNCCCNTNC CCTNNTNCNC TNCTNNTACC NTNCTNCTCC 1080

NTCTTCCTTC C 1091

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1165 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

NCNNNTTATG ATTACGCCNA CGNNCAATTA ACCTCACTAA AGGGAACAAA AGCTGGAGCT 60

CCACCGCGGT GGCGGCCGCT CTAGAGCTCG CGGCCGCGAG CTCAATTAAC CCTCACTAAA 120

GGGAGTCGAC TCGATCAGAC TGTTACTGTG TCTATGTAGA AAGAAGTAGA CATAAGAGAT 180

TCCATTTTGT TCTGTACTAA GAAAAATTCT TCTGCCTTGA GATGCTGTTA ATCTGTAACC 240

CTAGCCCCAA CCCTGTGCTC ACAGAGACAT GTGCTGTGTT GACTCAAGGT TCAATGGATT 300

TAGGGCTATG CTTTGTTAAA AAAGTGCTTG AAGATAATAT GCTTGTTAAA AGTCATCACC 360

ATTCTCTAAT CTCAAGTACC CAGGGACACA ATACACTGCG GAAGGCCGCA GGGACCTCTG 420

TCTAGGAAAG CCAGGTATTG TCCAAGATTT CTCCCCATGT GATAGCCTGA GATATGGCCT 480

CATGGGAAGG GTAAGACCTG ACTGTCCCCC AGCCCGACAT CCCCCAGCCC GACATCCCCC 540

AGCCCGACAC CCGAAAAGGG TCTGTGCTGA GGAAGATTAN TAAAAGAGGA AGGCTCTTTG 600

CATTGAAGTA AGAAGAAGGC TCTGTCTCCT GCTCGTCCCT GGGCAATAAA ATGTCTTGGT 660

GTTAAACCCG AATGTATGTT CTACTTACTG AGAATAGGAG AAAACATCCT TAGGGCTGGA 720

GGTGAGACAC CCTGGCGGCA TACTGCTCTT TAATGCACGA GATGTTTGTN TAATTGCCAT 780

CCAGGGCCAN CCCCTTTCCT TAACTTTTTA TGANACAAAA ACTTTGTTCN CTTTTCCTGC 840

GAACCTCTCC CCCTATTANC CTATTGGCCT GCCCATCCCC TCCCCAAANG GTGAAAANAT 900

GTTCNTAAAT NCGAGGGAAT CCAAAACNTT TTCCCGTTGG TCCCCTTTCC AACCCCGTCC 960

CTGGGCCNNT TTCCTCCCCA ACNTGTCCCG GNTCCTTCNT TCCCNCCCCC TTCCCNGANA 1020

AAAAACCCCG TNTGANGGNG CCCCCTCAAA TTATAACCTT TCCNAAACAA ANNGGTTCNA 1080

AGGTGGTTTG NTTCCGGTGC GGCTGGCCTT GAGGTCCCCC CTNCACCCCA ATTTGGAANC 1140

CNGTTTTTTT TATTGCCCNN TCCCC 1165

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1177 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

NCCNTTTAGA TGTTGACAAN NTAAACAAGC NGCTCAGGCA GCTGAAAAAA GCCACTGATA 60

AAGCATCCTG GAGTATCAGA GTTTACTGTT AGATCAGCCT CATTTGACTT CCCCTCCCAC 120

ATGGTGTTTA AATCCAGCTA CACTACTTCC TGACTCAAAC TCCACTATTC CTGTTCATGA 180

CTGTCAGGAA CTGTTGGAAA CTACTGAAAC TGGCCGACCT GATCTTCAAA ATGTGCCCCT 240

AGGAAAGGTG GATGCCACCG TGTTCACAGA CAGTACCNCC TTCCTCGAGA AGGGACTACG 300

AGGGGCCGGT GCANCTGTTA CCAAGGAGAC TNATGTGTTG TGGGCTCAGG CTTTACCANC 360

AAACACCTCA NCNCNNAAGG CTGAATTGAT CGCCCTCACT CAGGCTCTCG GATGGGGTAA 420

GGGATATTAA CGTTAACACT GACAGCAGGT ACGCCTTTGC TACTGTGCAT GTACGTGGAG 480

CCATCTACCA GGAGCGTGGG CTACTCACTC GGCAGGTGGC TGTNATCCAC TGTAAANGGA 540

CATCAAAAGG AAAACNNGGC TGTTGCCCGT GGTAACCANA AANCTGATCN NCAGCTCNAA 600

GATGCTGTGT TGACTTTCAC TCNCNCCTCT TAAACTTGCT GCCCACANTC TCCTTTCCCA 660

ACCAGATCTG CCTGACAATC CCCATACTCA AAAAAAAAAN AANACTGGCC CCGAACCCNA 720

ACCAATAAAA ACGGGGANGG TNGGTNGANC NNCCTGACCC AAAAATAATG GATCCCCCGG 780

GCTGCAGGAA TTCAATTCAN CCTTATCNAT ACCCCCAACN NGGNGGGGGG GGCCNGTNCC 840

CATTNCCCCT NTATTNATTC TTTNNCCCCC CCCCCGGCNT CCTTTTTNAA CTCGTGAAAG 900

GGAAAACCTG NCTTACCAAN TTATCNCCTG GACCNTCCCC TTCCNCGGTN GNTTANAAAA 960

AAAAGCCCNC ANTCCCNTCC NAAATTTGCA CNGAAAGGNA AGGAATTTAA CCTTTATTTT 1020

TTNNTCCTTT ANTTTGTNNN CCCCCTTTTA CCCAGGCGAA CNGCCATCNT TTAANAAAAA 1080

AAANAGAANG TTTATTTTTC CTTNGAACCA TCCCAATANA AANCACCCGC NGGGGAACGG 1140

GGNGGNAGGC CNCTCACCCC CTTTNTGTNG GNGGGNC 1177

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1146 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

NCCNNTTNNT GATGTTGTCT TTTTGGCCTC TCTTTGGATA CTTTCCCTCT CTTCAGAGGT 60

GAAAAGGGTC AAAAGGAGCT GTTGACAGTC ATCCCAGGTG GGCCAATGTG TCCAGAGTAC 120

AGACTCCATC AGTGAGGTCA AAGCCTGGGG CTTTTCAGAG AAGGGAGGAT TATGGGTTTT 180

CCAATTATAC AAGTCAGAAG TAGAAAGAAG GGACATAAAC CAGGAAGGGG GTGGAGCACT 240

CATCACCCAG AGGGACTTGT GCCTCTCTCA GTGGTAGTAG AGGGGCTACT TCCTCCCACC 300

ACGGTTGCAA CCAAGAGGCA ATGGGTGATG AGCCTACAGG GGACATANCC GAGGAGACAT 360

GGGATGACCC TAAGGGAGTA GGCTGGTTTT AAGGCGGTGG GACTGGGTGA GGGAAACTCT 420

CCTCTTCTTC AGAGAGAAGC AGTACAGGGC GAGCTGAACC GGCTGAAGGT CGAGGCGAAA 480

ACACGGTCTG GCTCAGGAAG ACCTTGGAAG TAAAATTATG AATGGTGCAT GAATGGAGCC 540

ATGGAAGGGG TGCTCCTGAC CAAACTCAGC CATTGATCAA TGTTAGGGAA ACTGATCAGG 600

GAAGCCGGGA ATTTCATTAA CAACCCGCCA CACAGCTTGA ACATTGTGAG GTTCAGTGAC 660

CCTTCAAGGG GCCACTCCAC TCCAACTTTG GCCATTCTAC TTTGCNAAAT TTCCAAAACT 720

TCCTTTTTTA AGGCCGAATC CNTANTCCCT NAAAAACNAA AAAAAATCTG CNCCTATTCT 780

GGAAAAGGCC CANCCCTTAC CAGGCTGGAA GAAATTTTNC CTTTTTTTTT TTTTTGAAGG 840

CNTTTNTTAA ATTGAACCTN AATTCNCCCC CCCAAAAAAA AACCCNCCNG GGGGGCGGAT 900

TTCCAAAAAC NAATTCCCTT ACCAAAAAAC AAAAACCCNC CCTTNTTCCC TTCCNCCCTN 960

TTCTTTTAAT TAGGGAGAGA TNAAGCCCCC CAATTTCCNG GNCTNGATNN GTTTCCCCCC 1020

CCCCCATTTT CCNAAACTTT TTCCCANCNA GGAANCCNCC CTTTTTTTNG GTCNGATTNA 1080

NCAACCTTCC AAACCATTTT TCCNNAAAAA NTTTGNTNGG NGGGAAAAAN ACCTNNTTTT 1140

ATAGAN 1146

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 545 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTTCATTGGG TACGGGCCCC CTCGACCTCG ACGGTATCGA TAAGCTTGAT ATCGAATTCC 60

TGCAGCCCGG GGGATCCACT AGTTCTAGAG TCAGGAAGAA CCACCAACCT TCCTGATTTT 120

TATTGGCTCT GAGTTCTGAG GCCAGTTTTC TTCTTCTGTT GAGTATGCGG GATTGTCAGG 180

CAGATCTGGC TGTGGAAAGG AGACTGTGGG CAGCAAGTTT AGAGGCGTGA CTGAAAGTCA 240

CACTGCATCT TGAGCTGCTG AATCAGCTTT CTGGTTACCA CGGGCAACAG CCGTGTTTTC 300

CTTTTGATGT CCTTTACAGT GGATTACAGC CACCTGCTGA GGTGAGTAGC CCACGCTCCT 360

GGTAGATGGC TCCACGTACA TGCACAGTAG CAAAGGCGTA CCTGCTGTCA GTGTTAACGT 420

TAATATCCTT ACCCCATCGG AGAGCCTGAG TGAGGGCGAT CAATTCAGCC CTTTTGTGCT 480

GAGGTGTTTG CTGGTTAAGC CCTGAACCCA CAACACATCT GTCTCCATGG TAACAGCTGC 540

ACCGG 545

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9388 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCTCGCGGCC GCGAGCTCAA TTAACCCTCA CTAAAGGGAG TCGACTCGAT CAGACTGTTA 60

CTGTGTCTAT GTAGAAAGAA GTAGACATAA GAGATTCCAT TTTGTTCTGT ACTAAGAAAA 120

ATTCTTCTGC CTTGAGATGC TGTTAATCTG TAACCCTAGC CCCAACCCTG TGCTCACAGA 180

GACATGTGCT GTGTTGACTC AAGGTTCAAT GGATTTAGGG CTATGCTTTG TTAAAAAAGT 240

GCTTGAAGAT AATATGCTTG TTAAAAGTCA TCACCATTCT CTAATCTCAA GTACCCAGGG 300

ACACAATACA CTGCGGAAGG CCGCAGGGAC CTCTGTCTAG GAAAGCCAGG TATTGTCCAA 360

GATTTCTCCC CATGTGATAG CCTGAGATAT GGCCTCATGG GAAGGGTAAG ACCTGACTGT 420

CCCCCAGCCC GACATCCCCC AGCCCGACAT CCCCCAGCCC GACACCCGAA AAGGGTCTGT 480

GCTGAGGAGG ATTAGTAAAA GAGGAAGGCC TCTTTGCAGT TGAGGTAAGA GGAAGGCATC 540

TGTCTCCTGC TCGTCCCTGG GCAATAGAAT GTCTTGGTGT AAAACCCGAT TGTATGTTCT 600

ACTTACTGAG ATAGGAGAAA ACATCCTTAG GGCTGGAGGT GAGACACGCT GGCGGCAATA 660

CTGCTCTTTA ATGCACCGAG ATGTTTGTAT AAGTGCACAT CAAGGCACAG CACCTTTCCT 720

TAAACTTATT TATGACACAG AGACCTTTGT TCACGTTTTC CTGCTGACCC TCTCCCCACT 780

ATTACCCTAT TGGCCTGCCA CATCCCCCTC TCCGAGATGG TAGAGATAAT GATCAATAAA 840

TACTGAGGGA ACTCAGAGAC CAGTGTCCCT GTAGGTCCTC CGTGTGCTGA GCGCCGGTCC 900

CTTGGGCTCA CTTTTCTTTC TCTATACTTT GTCTCTGTGT CTCTTTCTTT TCTCAGTCTC 960

TCGTTCCACC TGACGAGAAA TACCCACAGG TGTGGAGGGG CAGGCCACCC CTTCAATAAT 1020

TTACTAGCCT GTTCGCTGAC AACAAGACTG GTGGTGCAGA AGGTTGGGTC TTGGTGTTCA 1080

CCGGGTGGCA GGCATGGGCC AGGTGGGAGG GTCTCCAGCG CCTGGTGCAA ATCTCCAAGA 1140

AAGTGCAGGA AACAGCACCA AGGGTGATTG TAAATTTTGA TTTGGCGCGG CAGGTAGCCA 1200

TTCCAGCGCA AAAATGCGCA GGAAAGCTTT TGCTGTGCTT GTAGGCAGGT AGGCCCCAAG 1260

CACTTCTTAT TGGCTAATGT GGAGGGAACC TGCACATCCA TTGGCTGAAA TCTCCGTCTA 1320

TTTGAGGCTG ACTGAGCGCG TTCCTTTCTT CTGTGTTGCC TGGAAACGGA CTGTCTGCCT 1380

AGTAACATCT GATCACGTTT CCCATTGGCC GCCGTTTCCG GAAGCCCGCC CTCCCATTTC 1440

CGGAAGCCTG GCGCAAGGTT GGTCTGCAGG TGGCCTCCAG GTGCAAAGTG GGAAGTGTGA 1500

GTCCTCAGTC TTGGGCTATT CGGCCACGTG CCTGCCGGAC ATGGGACGCT GGAGGGTCAG 1560

CAGCGTGGAG TCCTGGCCTT TTGCGTCCAC GGGTGGGAAA TTGGCCATTG CCACGGCGGG 1620

AACTGGGACT CAGGCTGCCC CCCGGCCGTT TCTCATCCGT CCACCGGACT CGTGGGCGCT 1680

CGCACTGGCG CTGATGTAGT TTCCTGACCT CTGACCCGTA TTGTCTCCAG ATTAAAGGTA 1740

AAAACGGGGC TTTTTCAGCC CACTCGGGTA AAACGCCTTT TGATTTCTAG GCAGGTGTTT 1800

TGTTGCACGC CTGGGAGGGA GTGACCCGCA GGTTGAGGTT TATTAAAATA CATTCCTGGT 1860

TTATGTTATG TTTATAATAA AGCACCCCAA CCTTTACAAA ATCTCACTTT TTGCCAGTTG 1920

TATTATTTAG TGGACTGTCT CTGATAAGGA CAGCCAGTTA AAATGGAATT TTGTTGTTGC 1980

TAATTAAACC AATTTTTAGT TTTGGTGTTT GTCCTAATAG CAACAACTTC TCAGGCTTTA 2040

TAAAACCATA TTTCTTGGGG GAAATTTCTG TGTAAGGCAC AGCGAGTTAG TTTGGAATTG 2100

TTTTAAAGGA AGTAAGTTCC TGGTTTTGAT ATCTTAGTAG TGTAATGCCC AACCTGGTTT 2160

TTACTAACCC TGTTTTTAGA CTCTCCCTTT CCTTAAATCA CCTAGCCTTG TTTCCACCTG 2220

AATTGACTCT CCCTTAGCTA AGAGCGCCAG ATGGACTCCA TCTTGGCTCT TTCACTGGCA 2280

GCCCCTTCCT CAAGGACTTA ACTTGTGCAA GCTGACTCCC AGCACATCCA AGAATGCAAT 2340

TAACTGTTAA GATACTGTGG CAAGCTATAT CCGCAGTTCC GAGGAATTCA TCCGATTGAT 2400

TATGCCCAAA AGCCCCGCGT CTATCACCTT GTAATAATCT TAAAGCCCCT GCACCTGGAA 2460

CTATTAACTT TCCTGTAACC ATTTATCCTT TTAACTTTTT TGCTTACTTT ATTTCTGTAA 2520

AATTGTTTTA ACTAGACCTC CCCTCCCCTT TCTAAACCAA AGTATAAAAG AAGATCTAGC 2580

CCCTTCTTCA GAGCGGAGAG AATTTTGAGC ATTAGCCATC TCTTGGCGGC CAGCTAAATA 2640

AATGGACTTT TAATTTGTCT CAAAGTGTGG CGTTTTCTCT AACTCGCTCA GGTACGACAT 2700

TTGGAGGCCC CAGCGAGAAA CGTCACCGGG AGAAACGTCA CCGGGCGAGA GCCGGGCCCG 2760

CTGTGTGCTC CCCCGGAAGG ACAGCCAGCT TGTAGGGGGG AGTGCCACCT GAAAAAAAAA 2820

TTTCCAGGTC CCCAAAGGGT GACCGTCTTC CGGAGGACAG CGGATCGACT ACCATGCGGG 2880

TGCCCACCAA AATTCCACCT CTGAGTCCTC AACTGCTGAC CCCGGGGTCA GGTAGGTCAG 2940

ATTTGACTTT GGTTCTGGCA GAGGGAAGCG ACCCTGATGA GGGTGTCCCT CTTTTGACTC 3000

TGCCCATTTC TCTAGGATGC TAGAGGGTAG AGCCCTGGTT TTCTGTTAGA CGCCTCTGTG 3060

TCTCTGTCTG GGAGGGAAGT GGCCCTGACA GGGGCCATCC CTTGAGTCAG TCCACATCCC 3120

AGGATGCTGG GGGACTGAGT CCTGGTTTCT GGCAGACTGG TCTCTCTCTC TCTCTTTTTC 3180

TATCTCTAAT CTTTCCTTGT TCAGGTTTCT TGGAGAATCT CTGGGAAAGA AAAAAGAAAA 3240

ACTGTTATAA ACTCTGTGTG AATGGTGAAT GAATGGGGGA GGACAAGGGC TTGCGCTTGT 3300

CCTCCAGTTT GTAGCTCCAC GGCGAAAGCT ACGGAGTTCA AGTGGGCCCT CACCTGCGGT 3360

TCCGTGGCGA CCTCATAAGG CTTAAGGCAG CATCCGGCAT AGCTCGATCC GAGCCGGGGG 3420

TTTATACCGG CCTGTCAATG CTAAGAGGAG CCCAAGTCCC CTAAGGGGGA GCGGCCAGGC 3480

GGGCATCTGA CTGATCCCAT CACGGGACCC CCTCCCCTTG TTTGTCTAAA AAAAAAAAAA 3540

GAAGAAACTG TCATAACTGT TTACATGCCC TAGGGTCAAC TGTTTGTTTT ATGTTTATTG 3600

TTCTGTTCGG TGTCTATTGT CTTGTTTAGT GGTTGTCAAG GTTTTGCATG TCAGGACGTC 3660

GATATTGCCC AAGACGTCTG GGTAAGAACT TCTGCAAGGT CCTTAGTGCT GATTTTTTGT 3720

CACAGGAGGT TAAATTTCTC ATCAATCATT TAGGCTGGCC ACCACAGTCC TGTCTTTTCT 3780

GCCAGAAGCA AGTCAGGTGT TGTTACGGGA ATGAGTGTAA AAAAACATTC GCCTGATTGG 3840

GATTTCTGGC ACCATGATGG TTGTATTTAG ATTGTCATAC CCCACATCCA GGTTGATTGG 3900

ACCTCCTCTA AACTAAACTG GTGGTGGGTT CAAAACAGCC ACCCTGCAGA TTTCCTTGCT 3960

CACCTCTTTG GTCATTCTGT AACTTTTCCT GTGCCCTTAA ATAGCACACT GTGTAGGGAA 4020

ACCTACCCTC GTACTGCTTT ACTTCGTTTA GATTCTTACT CTGTTCCTCT GTGGCTACTC 4080

TCCCATCTTA AAAACGATCC AAGTGGTCCT TTTCCTCCTC CCTGCCCCCT ACCCCACACA 4140

TCTCGTTTTC CAGTGCGACA GCAAGTTCAG CGTCTCCAGG ACTTGGCTCT GCTCTCACTC 4200

CTTGAACCCT TAAAAGAAAA AGCTGGGTTT GAGCTATTTG CCTTTGAGTC ATGGAGACAC 4260

AAAAGGTATT TAGGGTACAG ATCTAGAAGA AGAGAGAGAA CACCTAGATC CAACTGACCC 4320

AGGAGATCTC GGGCTGGCCT CTAGTCCTCC TCCCTCAATC TTAAAGCTAC AGTGATGTGG 4380

CAAGTGGTAT TTAGCTGTTG TGGTTTTTCT GCTCTTTCTG GTCATGTTGA TTCTGTTCTT 4440

TCGATACTCC AGCCCCCCAG GGAGTGAGTT TCTCTGTCTG TGCTGGGTTT GATATCTATG 4500

TTCAAATCTT ATTAAATTGC CTTCAAAAAA AAAAAAAAAA GGGAAACACT TCCTCCCAGC 4560

CTTGTAAGGG TTGGAGCCCT CTCCAGTATA TGCTGCAGAA TTTTTCTCTC GGTTTCTCAG 4620

AGGATTATGG AGTCCGCCTT AAAAAAGGCA AGCTCTGGAC ACTCTGCAAA GTAGAATGGC 4680

CAAAGTTTGG AGTTGAGTGG CCCCTTGAAG GGTCACTGAA CCTCACAATT GTTCAAGCTG 4740

TGTGGCGGGT TGTTACTGAA ACTCCCGGCC TCCCTGATCA GTTTCCCTAC ATTGATCAAT 4800

GGCTGAGTTT GGTCAGGAGC ACCCCTTCCA TGGCTCCACT CATGCACCAT TCATAATTTT 4860

ACCTCCAAGG TCCTCCTGAG CCAGACCGTG TTTTCGCCTC GACCCTCAGC CGGTTCAGCT 4920

CGCCCTGTAC TGCCTCTCTC TGAAGAAGAG GAGAGTCTCC CTCACCCAGT CCCACCGCCT 4980

TAAAACCAGC CTACTCCCTT AGGGTCATCC CATGTCTCCT CGGCTATGTC CCCTGTAGGC 5040

TCATCACCCA TTGCCTCTTG GTTGCAACCG TGGTGGGAGG AAGTAGCCCC TCTACTACCA 5100

CTGAGAGAGG CACAAGTCCC TCTGGGTGAT GAGTGCTCCA CCCCCTTCCT GGTTTATGTC 5160

CCTTCTTTCT ACTTCTGACT TGTATAATTG GAAAACCCAT AATCCTCCCT TCTCTGAAAA 5220

GCCCCAGGCT TTGACCTCAC TGATGGAGTC TGTACTCTGG ACACATTGGC CCACCTGGGA 5280

TGACTGTCAA CAGCTCCTTT TGACCCTTTT CACCTCTGAA GAGAGGGAAA GTATCCAAAG 5340

AGAGGCCAAA AAGTACAACC TCACATCAAC CAATAGGCCG GAGGAGGAAG CTAGAGGAAT 5400

AGTGATTAGA GACCCAATTG GGACCTAATT GGGACCCAAA TTTCTCAAGT GGAGGGAGAA 5460

CTTTTGACGA TTTCCACCGG TATCTCCTCG TGGGTATTCA GGGAGCTGCT CAGAAACCTA 5520

TAAACTTGTC TAAGGCGACT GAAGTCGTCC AGGGGCATGA TGAGTCACCA GGAGTGTTTT 5580

TAGAGCACCT CCAGGAGGCT TATCGGATTT ACACCCCTTT TGACCTGGCA GCCCCCGAAA 5640

ATAGCCATGC TCTTAATTTG GCATTTGTGG CTCAGGCAGC CCCAGATAGT AAAAGGAAAC 5700

TCCAAAAACT AGAGGGATTT TGCTGGAATG AATACCAGTC AGCTTTTAGA GATAGCCTAA 5760

AAGGTTTTTG ACAGTCAAGA GGTTGAAAAA CAAAAACAAG CAGCTCAGGC AGCTGAAAAA 5820

AGCCACTGAT AAAGCATCCT GGAGTATCAG AGTTTACTGT TAGATCAGCC TCATTTGACT 5880

TCCCCTCCCA CATGGTGTTT AAATCCAGCT ACACTACTTC CTGACTCAAA CTCCACTATT 5940

CCTGTTCATG ACTGTCAGGA ACTGTTGGAA ACTACTGAAA CTGGCCGACC TGATCTTCAA 6000

AATGTGCCCC TAGGAAAGGT GGATGCCACC GTGTTCACAG ACAGTAGCAG CTTCCTCGAG 6060

AAGGGACTAC GAAAGGCCGG TGCAGCTGTT ACCATGGAGA CAGATGTGTT GTGGGCTCAG 6120

GCTTTACCAG CAAACACCTC AGCACAAAAG GCTGAATTGA TCGCCCTCAC TCAGGCTCTC 6180

CGATGGGGTA AGGATATTAA CGTTAACACT GACAGCAGGT ACGCCTTTGC TACTGTGCAT 6240

GTACGTGGAG CCATCTACCA GGAGCGTGGG CTACTCACCT CAGCAGGTGG CTGTAATCCA 6300

CTGTAAAGGA CATCAAAAGG AAAACACGGC TGTTGCCCGT GGTAACCAGA AAGCTGATTC 6360

AGCAGCTCAA GATGCAGTGT GACTTTCAGT CACGCCTCTA AACTTGCTGC CCACAGTCTC 6420

CTTTCCACAG CCAGATCTGC CTGACAATCC CGCATACTCA ACAGAAGAAG AAAACTGGCC 6480

TCAGAACTCA GAGCCAATAA AAATCAGGAA GGTTGGTGGA TTCTTCCTGA CTCTAGAATC 6540

TTCATACCCC GAACTCTTGG GAAAACTTTA ATCAGTCACC TACAGTCTAC CACCCATTTA 6600

GGAGGAGCAA AGCTACCTCA GCTCCTCCGG AGCCGTTTTA AGATCCCCCA TCTTCAAAGC 6660

CTAACAGATC AAGCAGCTCT CCGGTGCACA ACCTGCGCCC AGGTAAATGC CAAAAAAGGT 6720

CCTAAACCCA GCCCAGGCCA CCGTCTCCAA GAAAACTCAC CAGGAGAAAA GTGGGAAATT 6780

GACTTTACAG AAGTAAAACC ACACCGGGCT GGGTACAAAT ACCTTCTAGT ACTGGTAGAC 6840

ACCTTCTCTG GATGGACTGA AGCATTTGCT ACCAAAAACG AAACTGTCAA TATGGTAGTT 6900

AAGTTTTTAC TCAATGAAAT CATCCCTCGA CGTGGGCTGC CTGTTGCCAT AGGGTCTGAT 6960

AATGGACCGG CCTTCGCCTT GTCTATAGTT TAGTCAGTCA GTAAGGCGTT AAACATTCAA 7020

TGGAAGCTCC ATTGTGCCTA TCGACCCCAG AGCTCTGGGC AAGTAGAACG CATGAACTGC 7080

ACCCTAAAAA ACACTCTTAC AAAATTAATC TTAGAAACCG GTGTAAATTG TGTAAGTCTC 7140

CTTCCTTTAG CCCTACTTAG AGTAAGGTGC ACCCCTTACT GGGCTGGGTT CTTACCTTTT 7200

GAAATCATGT ATGGGAGGGC GCTGCCTATC TTGCCTAAGC TAAGAGATGC CCAATTGGCA 7260

AAAATATCAC AAACTAATTT ATTACAGTAC CTACAGTCTC CCCAACAGGT ACAAGATATC 7320

ATCCTGCCAC TTGTTCGAGG AACCCATCCC AATCCAATTC CTGAACAGAC AGGGCCCTGC 7380

CATTCATTCC CGCCAGGTGA CCTGTTGTTT GTTAAAAAGT TCCAGAGAGA AGGACTCCCT 7440

CCTGCTTGGA AGAGACCTCA CACCGTCATC ACGATGCCAA CGGCTCTGAA GGTGGATGGC 7500

ATTCCTGCGT GGATTCATCA CTCCCGCATC AAAAAGGCCA ACGGAGCCCA ACTAGAAACA 7560

TGGGTCCCCA GGGCTGGGTC AGGCCCCTTA AAACTGCACC TAAGTTGGGT GAAGCCATTA 7620

GATTAATTCT TTTTCTTAAT TTTGTAAAAC AATGCATAGC TTCTGTCAAA CTTATGTATC 7680

TTAAGACTCA ATATAACCCC CTTGTTATAA CTGAGGAATC AATGATTTGA TTCCCCAAAA 7740

ACACAAGTGG GGAATGTAGT GTCCAACCTG GTTTTTACTA ACCCTGTTTT TAGACTCTCC 7800

CTTTCCTTTA ATCACTCAGC CTTGTTTCCA CCTGAATTGA CTCTCCCTTA GCTAAGAGCG 7860

CCAGATGGAC TCCATCTTGG CTCTTTCACT GGCAGCCGCT TCCTCAAGGA CTTAACTTGT 7920

GCAAGCTGAC TCCCAGCACA TCCAAGAATG CAATTAACTG ATAAGATACT GTGGCAAGCT 7980

ATATCCGCAG TTCCCAGGAA TTCGTCCAAT TGATTACACC CAAAAGCCCC GCGTCTATCA 8040

CCTTGTAATA ATCTTAAAGC CCCTGCACCT GGAACTATTA ACGTTCCTGT AACCATTTAT 8100

CCTTTTAACT TTTTTGCCTA CTTTATTTCT GTAAAATTGT TTTAACTAGA CCCCCCCTCT 8160

CCTTTCTAAA CCAAAGTATA AAAGCAAATC TAGCCCCTTC TTCAGGCCGA GAGAATTTCG 8220

AGCGTTAGCC GTCTCTTGGC CACCAGCTAA ATAAACGGAT TCTTCATGTG TCTCAAAGTG 8280

TGGCGTTTTC TCTAACTCGC TCAGGTACGA CCGTGGTAGT ATTTTCCCCA ACGTCTTATT 8340

TTTAGGGCAC GTATGTAGAG TAACTTTTAT GAAAGAAACC AGTTAAGGAG GTTTTGGGAT 8400

TTCCTTTATC AACTGTAATA CTGGTTTTGA TTATTTATTT ATTTATTTAT TTTTTTTGAG 8460

AAGGAGTTTC ACTCTTGTTG CCCAGGCTGG AGTGCAATGG TGCGATCTTG GCTCACTGCA 8520

ACTTCCGCCT CCCAGGTTCA AGCGATTCTC CTGCCTCAGC CTCGAGAGTA GCTGGGATTA 8580

TAGGCATGCG CCACCACACC CAGCTAATTT TGTATTTTTA GTAAAGATGG GGTTTCTTCA 8640

TGTTGGTCAA GCTGGTCTGG AACTCCCCGC CTCGGGTGAT CTGCCCGCCT CGGCCTCCGA 8700

AAGTGCTGGG ATTACAGGTG TGATCCACCA CACCCAGCCG ATTTATATGT ATATAAATCA 8760

CATTCCTCTA ACCAAAATGT AGTGTTTCCT TCCATCTTGA ATATAGGCTG TAGACCCCGT 8820

GGGTATGGGA CATTGTTAAC AGTGAGACCA CAGCAGTTTT TATGTCATCT GACAGCATCT 8880

CCAAATAGCC TTCATGGTTG TCACTGCTTC CCAAGACAAT TCCAAATAAC ACTTCCCAGT 8940

GATGACTTGC TACTTGCTAT TGTTACTTAA TGTGTTAAGG TGGCTGTTAC AGACACTATT 9000

AGTATGTCAG GAATTACACC AAAATTTAGT GGCTCAAACA ATCATTTTAT TATGTATGTG 9060

GATTCTCATG GTCAGGTCAG GATTTCAGAC AGGGCACAAG GGTAGCCCAC TTGTCTCTGT 9120

CTATGATGTC TGGCCTCAGC ACAGGAGACT CAACAGCTGG GGTCTGGGAC CATTTGGAGG 9180

CTTGTTCCCT CACATCTGAT ACCTGGCTTG GGATGTTGGA AGAGGGGGTG AGCTGAGACT 9240

GAGTGCCTAT ATGTAGTGTT TCCATATGGC CTTGACTTCC TTACAGCCTG GCAGCCTCAG 9300

GGTAGTCAGA ATTCTTAGGA GGCACAGGGC TCCAGGGCAG ATGCTGAGGG GTCTTTTATG 9360

AGGTAGCACA GCAAATCCAC CCAGGATC 9388

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3646 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGAAACACT TCCTCCCAGC CTTGTAAGGG TTGGAGCCCT CTCCAGTATA TGCTGCAGAA 60

TTTTTCTCTC GGTTTCTCAG AGGATTATGG AGTCCGCCTT AAAAAAGGCA AGCTCTGGAC 120

ACTCTGCAAA GTAGAATGGC CAAAGTTTGG AGTTGAGTGG CCCCTTGAAG GGTCACTGAA 180

CCTCACAATT GTTCAAGCTG TGTGGCGGGT TGTTACTGAA ACTCCCGGCC TCCCTGATCA 240

GTTTCCCTAC ATTGATCAAT GGCTGAGTTT GGTCAGGAGC ACCCCTTCCG TGGCTCCACT 300

CATGCACCAT TCATAATTTT ACCTCCAAGG TCCTCCTGAG CCAGACCGTG TTTTCGCCTC 360

GACCCTCAGC CGGTTCGGCT CGCCCTGTAC TGCCTCTCTC TGAAGAAGAG GAGAGTCTCC 420

CTCACCCAGT CCCACCGCCT TAAAACCAGC CTACTCCCTT AGGGTCATCC CATGTCTCCT 480

CGGCTATGTC CCCTGTAGGC TCATCACCCA TTGCCTCTTG GTTGCAACCG TGGTGGGAGG 540

AAGTAGCCCC TCTACTACCA CTGAGAGAGG CACAAGTCCC TCTGGGTGAT GAGTGCTCCA 600

CCCCCTTCCT GGTTTATGTC CCTTCTTTCT ACTTCTGACT TGTATAATTG GAAAACCCAT 660

AATCCTCCCT TCTCTGAAAA GCCCCAGGCT TTGACCTCAC TGATGGAGTC TGTACTCTGG 720

ACACATTGGC CCACCTGGGA TGACTGTCAA CAGCTCCTTT TGACCCTTTT CACCTCTGAA 780

GAGAGGGAAA GTATCCAAAG AGAGGCCAAA AAGTACAACC TCACATCAAC CAATAGGCCG 840

GAGGAGGAAG CTAGAGGAAT AGTGATTAGA GACCCAATTG GGACCTAATT GGGACCCAAA 900

TTTCTCAAGT GGAGGGAGAA CTTTTGACGA TTTCCACCGG TATCTCCTCG TGGGTATTCA 960

GGGAGCTGCT CAGAAACCTA TAAACTTGTC TAAGGCGACT GAAGTCGTCC AGGGGCATGA 1020

TGAGTCACCA GGAGTGTTTT TAGAGCACCT CCAGGAGGCT TATCAGATTT ACACCCCTTT 1080

TGACCTGGCA GCCCCCGAAA ATAGCCATGC TCTTAATTTG GCATTTGTGG CTCAGGCAGC 1140

CCCAGATAGT AAAAGGAAAC TCCAAAAACT AGAGGGATTT TGCTGGAATG AATACCAGTC 1200

AGCTTTTAGA GATAGCCTAA AAGGTTTTTG ACAGTCAAGA GGTTGAAAAA CAAAAACAAG 1260

CAGCTCAGGC AGCTGAAAAA AGCCACTGAT AAAGCATCCT GGAGTATCAG AGTTTACTGT 1320

TAGATCAGCC TCATTTGACT TCCCCTCCCA CATGGTGTTT AAATCCAGCT ACACTACTTC 1380

CTGACTCAAA CTCCACTATT CCTGTTCATG ACTGTCAGGA ACTGTTGGAA ACTACTGAAA 1440

CTGGCCGACC TGATCTTCAA AATGTGCCCC TAGGAAAGGT GGATGCCACC ATGTTCACAG 1500

ACAGTAGCAG CTTCCTCGAG AAGGGACTAC GAAAGGCCGG TGCAGCTGTT ACCATGGAGA 1560

CAGATGTGTT GTGGGCTCAG GCTTTACCAG CAAACACCTC AGCACAAAAG GCTGAATTGA 1620

TCGCCCTCAC TCAGGCTCTC CGATGGGGTA AGGATATTAA CGTTAACACT GACAGCAGGT 1680

ACGCCTTTGC TACTGTGCAT GTACGTGGAG CCATCTACCA GGAGCGTGGG CTACTCACCT 1740

CAGCAGGTGG CTGTAATCCA CTGTAAAGGA CATCAAAAGG AAAACACGGC TGTTGCCCGT 1800

GGTAACCAGA AAGCTGATTC AGCAGCTCAA GATGCAGTGT GACTTTCAGT CACGCCTCTA 1860

AACTTGCTGC CCACAGTCTC CTTTCCACAG CCAGATCTGC CTGACAATCC CGCATACTCA 1920

ACAGAAGAAG AAAACTGGCC TCAGAACTCA GAGCCAATAA AAATCAGGAA GGTTGGTGGA 1980

TTCTTCCTGA CTCTAGAATC TTCATACCCC GAACTCTTGG GAAAACTTTA ATCAGTCACC 2040

TACAGTCTAC CACCCATTTA GGAGGAGCAA AGCTACCTCA GCTCCTCCGG AGCCGTTTTA 2100

AGATCCCCCA TCTTCAAAGC CTAACAGATC AAGCAGCTCT CCGGTGCACA ACCTGCGCCC 2160

AGGTAAATGC CAAAAAAGGT CCTAAACCCA GCCCAGGCCA CCGTCTCCAA GAAAACTCAC 2220

CAGGAGAAAA GTGGGAAATT GACTTTACAG AAGTAAAACC ACACCGGGCT GGGTACAAAT 2280

ACCTTCTAGT ACTGGTAGAC ACCTTCTCTG GATGGACTGA AGCATTTGCT ACCAAAAACG 2340

AAACTGTCAA TATGGTAGTT AAGTTTTTAC TCAATGAAAT CATCCCTCGA CATGGGCTGC 2400

CTGTTTGCCA TAGGGTCTGA TAATGGACCG GCCTTCGCCT TGTCTATAGT TTAGTCAGTC 2460

AGTAAGGCGT TAAACATTCA ATGGAAGCTC CATTGTGCCT ATCGACCCCA GAGCTCTGGG 2520

CAAGTAGAAC GCATGAACTG CACCCTAAAA AACACTCTTA CAAAATTAAT CTTAGAAACC 2580

GGTGTAAATT GTGTAAGTCT CCTTCCTTTA GCCCTACTTA GAGTAAGGTG CACCCCTTAC 2640

TGGGCTGGGT TCTTACCTTT TGAAATCATG TATGGGAGGG TGCTGCCTAT CTTGCCTAAG 2700

CTAAGAGATG CCCAATTGGC AAAAATATCA CAAACTAATT TATTACAGTA CCTACAGTCT 2760

CCCCAACAGG TACAAGATAT CATCCTGCCA CTTGTTCGAG GAACCCATCC CAATCCAATT 2820

CCTGAACAGA CAGGGCCCTG CCATTCATTC CCGCCAGGTG ACCTGTTGTT TGTTAAAAAG 2880

TTCCAGAGAG AAGGACTCCC TCCTGCTTGG AAGAGACCTC ACACCGTCAT CACGATGCCA 2940

ACGGCTCTGA AGGTGGATGG CATTCCTGCG TGGATTCATC ACTCCCGCAT CAAAAAGGCC 3000

AACAGAGCCC AACTAGAAAC ATGGGTCCCC AGGGCTGGGT CAGGCCCCTT AAAACTGCAC 3060

CTAAGTTGGG TGAAGCCATT AGATTAATTC TTTTTCTTAA TTTTGTAAAA CAATGCATAG 3120

CTTCTGTCAA ACTTATGTAT CTTAAGACTC AATATAACCC CCTTGTTATA ACTGAGGAAT 3180

CAATGATTTG ATTCCCCCAA AAACACAAGT GGGGAATGTA GTGTCCAACC TGGTTTTTAC 3240

TAACCCTGTT TTTAGACTCT CCCTTTCCTT TAATCACTCA GCTTGTTTCC ACCTGAATTG 3300

ACTCTCCCTT AGCTAAGAGC GCCAGATGGA CTCCATCTTG GCTCTTTCAC TGGCAGCCGC 3360

TTCCTCAAGG ACTTAACTTG TGCAAGCTGA CTCCCAGCAC ATCCAAGAAT GCAATTAACT 3420

GATAAGATAC TGTGGCAAGC TATATCCGCA GTTCCCAGGA ATTCGTCCAA TTGATCACAG 3480

CCCCTCTACC CTTCAGCAAC CACCACCCTG ATCAGTCAGC AGCCATCAGC ACCGAGGCAA 3540

GGCCCTCCAC CAGCAAAAAG ATTCTGACTC ACTGAAGACT TGGATGATCA TTAGTATTTT 3600

TAGCAGTAAA GTTTTTTTTT CTTTTTCTTT CTTTTTTTCT CGTGCC 3646

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCTCAACCTC 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATGGCTATTT TCGGGGGCTG ACA 23

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCGGTATCTC CTCGTGGGTA TT 22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTTCAACCTC 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTGCCTGAGC CACAAATG 18

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCGGAGGAGG AAGCTAGAGG AATA 24

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTTTTTTTTT TTAG 14

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ser Ser Gly Gly Arg Thr Phe Asp Asp Phe His Arg Tyr Leu Leu Val
1 5 10 15

Gly Ile

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Gln Gly Ala Ala Gln Lys Pro Ile Asn Leu Ser Lys Xaa Ile Glu Val
1 5 10 15

Val Gln Gly His Asp Glu

20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ser Pro Gly Val Phe Leu Glu His Leu Gln Glu Ala Tyr Arg Ile Tyr
1 5 10 15

Thr Pro Phe Asp Leu Ser Ala
20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Tyr Leu Leu Val Gly Ile Gln Gly Ala
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Ala Ala Gln Lys Pro Ile Asn Leu
1 5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Asn Leu Ser Lys Xaa Ile Glu Val Val
1 5

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Glu Val Val Gln Gly His Asp Glu Ser
1 5

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

His Leu Gln Glu Ala Tyr Arg Ile Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Asn Leu Ala Phe Val Ala Gln Ala Ala
1 5

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Phe Val Ala Gln Ala Ala Pro Asp Ser
1 5

Claims

1. An isolated DNA molecule, comprising:

(a) a human endogenous retroviral sequence, wherein said retroviral sequence is preferentially expressed in a tumor tissue;

(b) a variant of said human endogenous retroviral sequence that contains one or more nucleotide substitutions, deletions, insertions and/or modifications at no more than 20% of the nucleotide positions, such that the antigenic and/or immunogenic properties of the polypeptide encoded by the human endogenous retroviral sequence are retained; or

(c) a nucleotide sequence encoding an epitope of a polypeptide encoded by at least one of the above sequences.

2. An isolated DNA molecule encoding an epitope of a polypeptide, wherein said polypeptide is encoded by:

(a) a nucleotide sequence transcribed from the sequence of SEQ ID NO:11; or

(b) a variant of said nucleotide sequence that contains one or more nucleotide substitutions, deletions, insertions and/or modifications at no more than 20% of the nucleotide positions, such that the antigenic and/or immunogenic properties of the polypeptide encoded by the nucleotide sequence are retained.

3. A recombinant expression vector comprising a DNA molecule according to claim 1 or claim 2.

4. A host cell transformed or transfected with an expression vector according to claim 3.

5. A polypeptide comprising an amino acid sequence encoded by a DNA molecule according to claim 1 or claim 2.

6. A monoclonal antibody that binds to a polypeptide according to claim 5.

7. A method for determining the presence of a cancer in a patient comprising detecting, within a biological sample obtained from a patient, a polypeptide according to claim 5, and therefrom determining the presence of cancer in the patient.

8. The method of claim 7 wherein the biological sample is a tumor sample.

9. The method of claim 7 wherein the step of detecting comprises contacting the biological sample with a monoclonal antibody according to claim 6.

10. The method of claim 7 wherein the polypeptide comprises an amino acid sequence encoded by a human endogenous retroviral sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 - SEQ ID NO:10 and SEQ ID NO:12.

11. A method for determining the presence of a cancer in a patient comprising detecting, within a biological sample obtained from a patient, an RNA molecule encoding a polypeptide according to claim 5, and therefrom determining the presence of cancer in the patient.

12. The method of claim 11 wherein the biological sample is a tumor sample.

13. The method of claim 11 wherein the step of detecting comprises:

(a) preparing cDNA from RNA molecules within the biological sample; and

(b) specifically amplifying cDNA molecules that are capable of encoding at least a portion of a polypeptide according to claim 5.

14. The method of claim 11 wherein the polypeptide comprises an amino acid sequence encoded by a human endogenous retroviral sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 - SEQ ID NO:10 and SEQ ID NO:12.

15. A polypeptide according to claim 5 for use within a method for detecting the presence of a cancer in a patient.

16. The polypeptide of claim 15 wherein the polypeptide comprises an amino acid sequence encoded by a human endogenous retroviral sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 - SEQ ID NO:10 and SEQ ID NO:12.

17. A method for monitoring the progression of a cancer in a patient, comprising:

(a) detecting an amount, in a biological sample obtained from a patient, of a polypeptide according to claim 5;

(b) subsequently repeating step (a); and

(c) comparing the amounts of polypeptide detected in steps (a) and (b), and therefrom monitoring the progression of cancer in the patient.

18. The method of claim 17 wherein the biological sample is a tumor sample.

19. The method of claim 17 wherein the step of detecting comprises contacting a portion of the biological sample with a monoclonal antibody according to claim 6.

20. The method of claim 17 wherein the polypeptide comprises an amino acid sequence encoded by a human endogenous retroviral sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 - SEQ ID NO:10 and SEQ ID NO:12.

21. A method for monitoring the progression of a cancer in a patient, comprising:

(a) detecting an amount, within a biological sample obtained from a patient, of an RNA molecule encoding a polypeptide according to claim 5;

(b) subsequently repeating step (a); and

(c) comparing the amounts of RNA molecules detected in steps (a) and (b), and therefrom monitoring the progression of cancer in the patient.

22. The method of claim 21 wherein the step of detecting comprises:

(a) preparing cDNA from RNA molecules within the biological sample; and

(b) specifically amplifying cDNA molecules that are capable of encoding at least a portion of a polypeptide according to claim 5.

23. The method of claim 21 wherein the polypeptide comprises an amino acid sequence encoded by a human endogenous retroviral sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3 - SEQ ID NO:10 and SEQ ID NO:12.

24. A pharmaceutical composition, comprising:

(a) a polypeptide according to claim 5; and

(b) a physiologically acceptable carrier.

25. A vaccine, comprising:
(a) a polypeptide according to claim 5; and
(b) an immune response enhancer.

26. A diagnostic kit comprising:
(a) one or more monoclonal antibodies according to claim 6; and
(b) a detection reagent.

27. The kit of claim 26 wherein the monoclonal antibody(s) are immobilized on a solid support.

28. A diagnostic kit comprising a first polymerase chain reaction primer and a second polymerase chain reaction primer, the first and second primers each comprising at least about 10 contiguous nucleotides of an RNA molecule encoding a polypeptide according to claim 5.

29. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe comprising at least about 15 contiguous nucleotides of a DNA molecule according to claim 1 or claim 2.

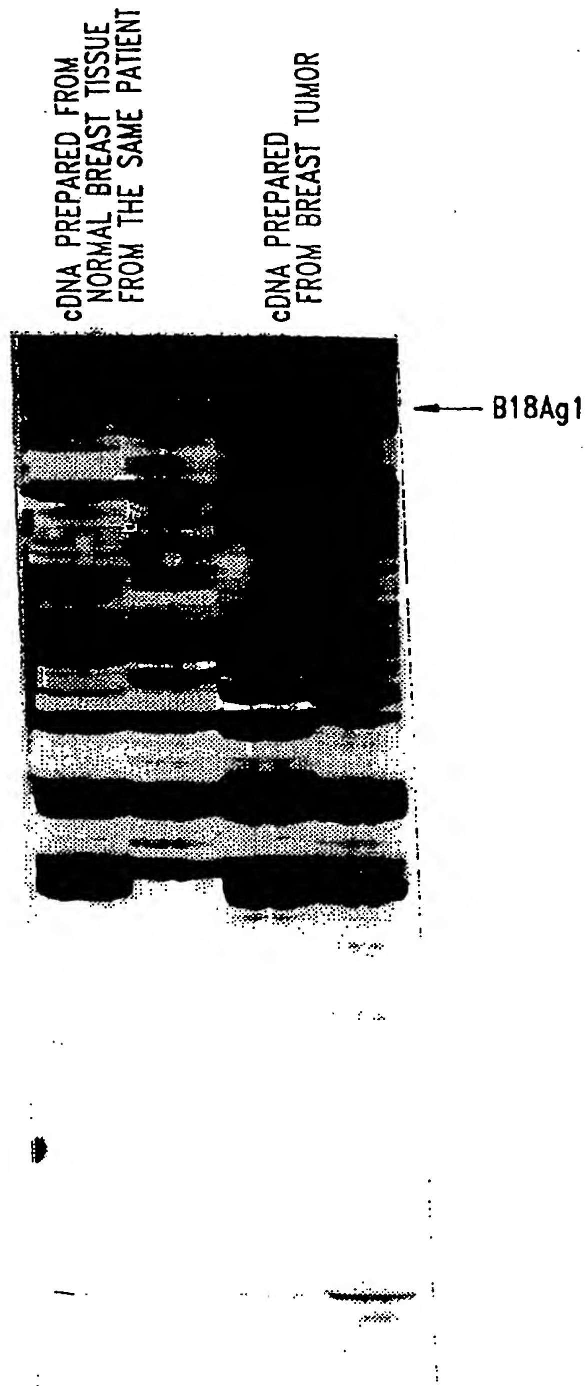

cDNA PREPARED FROM NORMAL BREAST TISSUE FROM THE SAME PATIENT
cDNA PREPARED FROM BREAST TUMOR
B18Ag1

Fig. 1

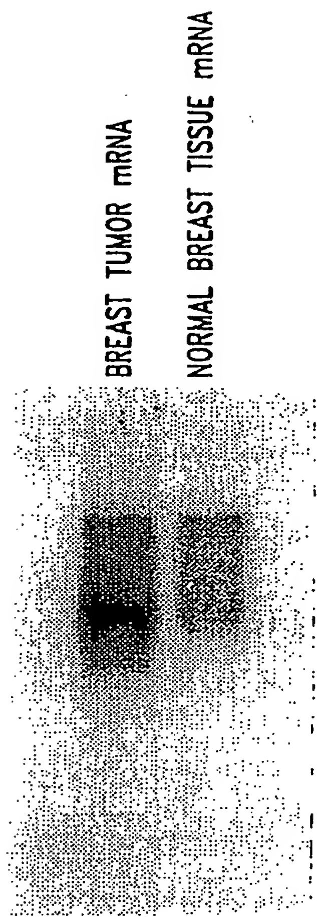
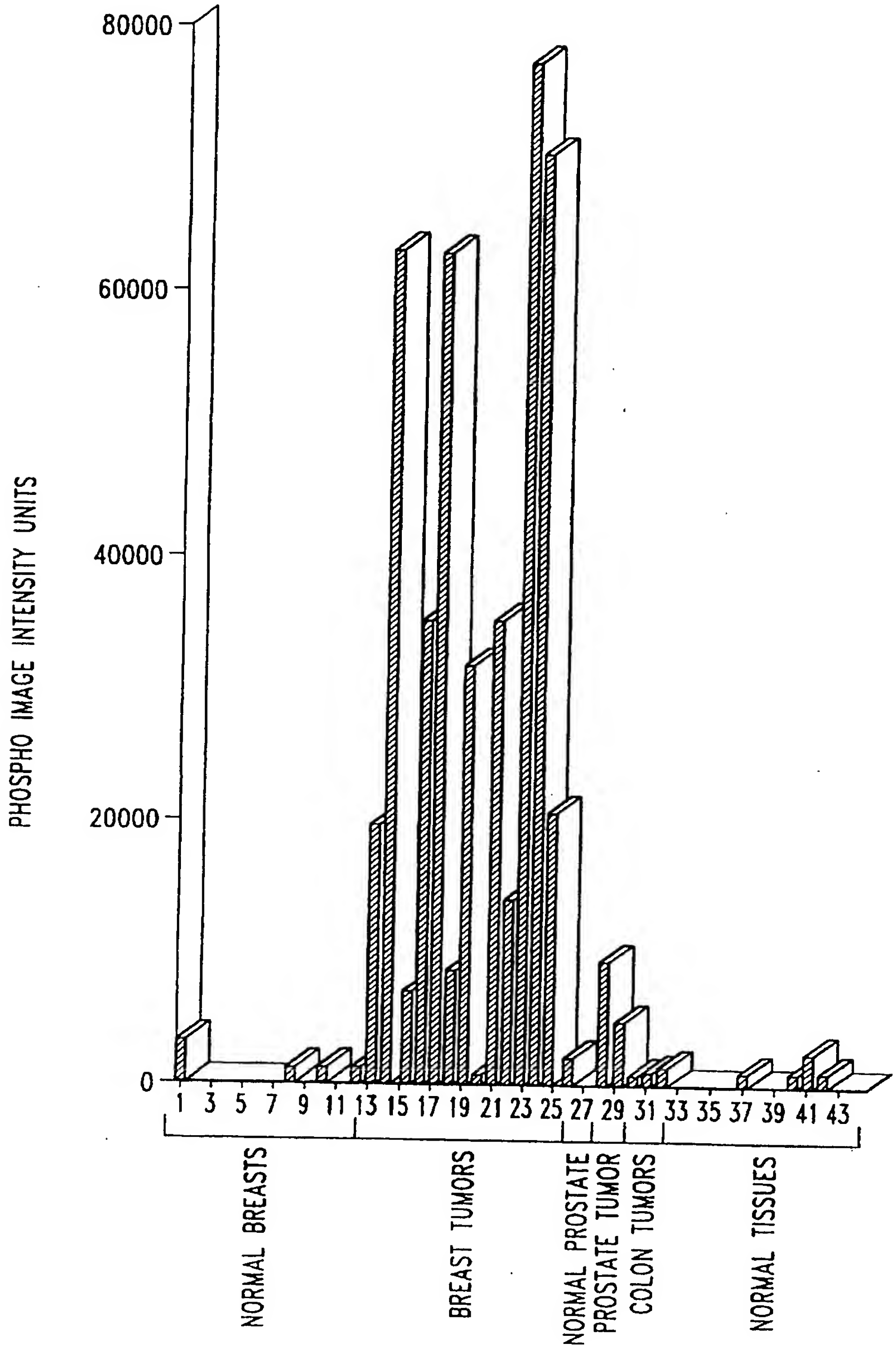


Fig. 2



PHOSPHO IMAGE INTENSITY UNITS
80000
60000
40000
20000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43
NORMAL BREASTS
BREAST TUMORS
NORMAL PROSTATE
PROSTATE TUMOR
COLON TUMORS
NORMAL TISSUES

Fig. 3

GENOMIC CLONE MAP

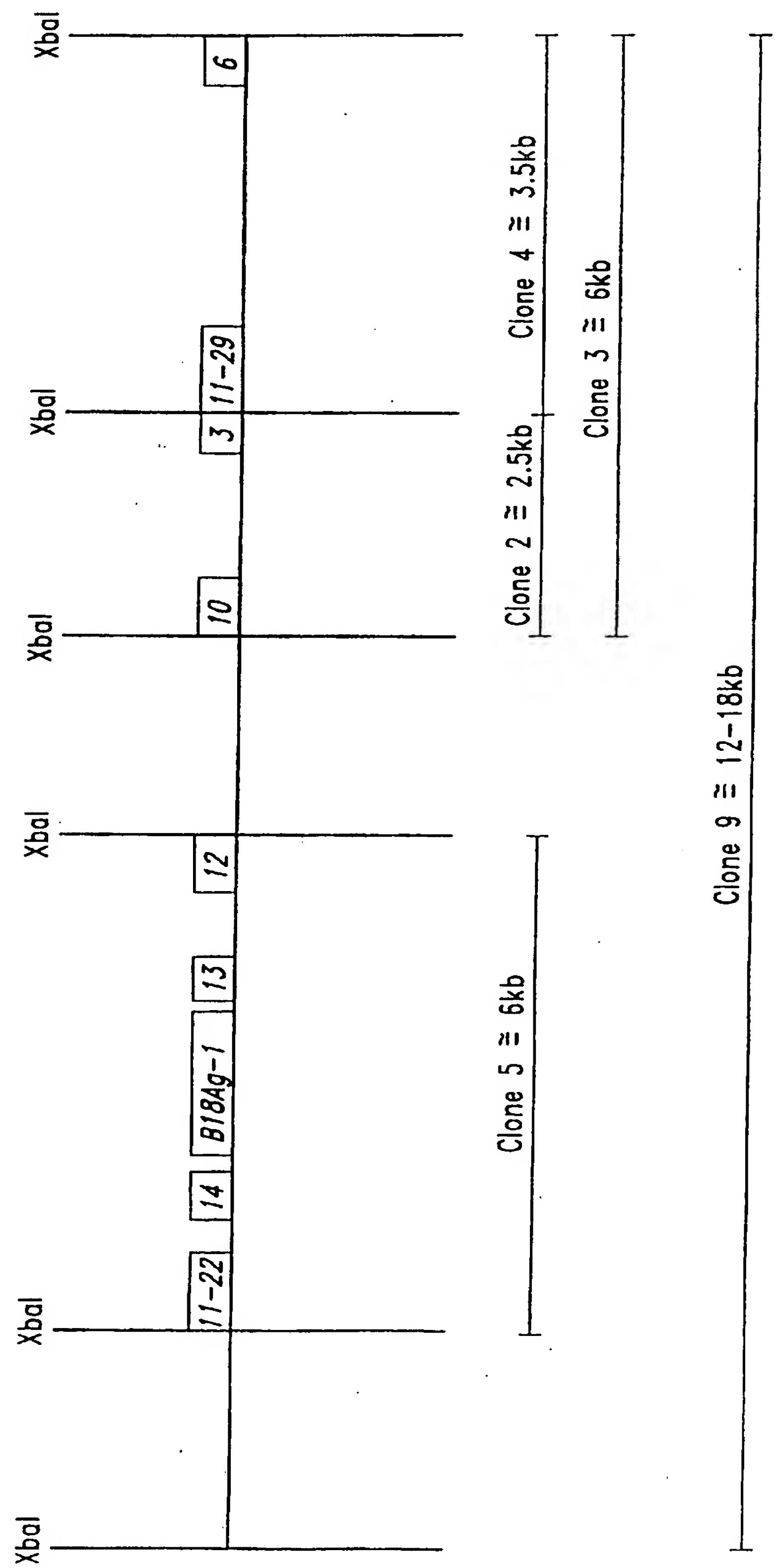


Fig. 4

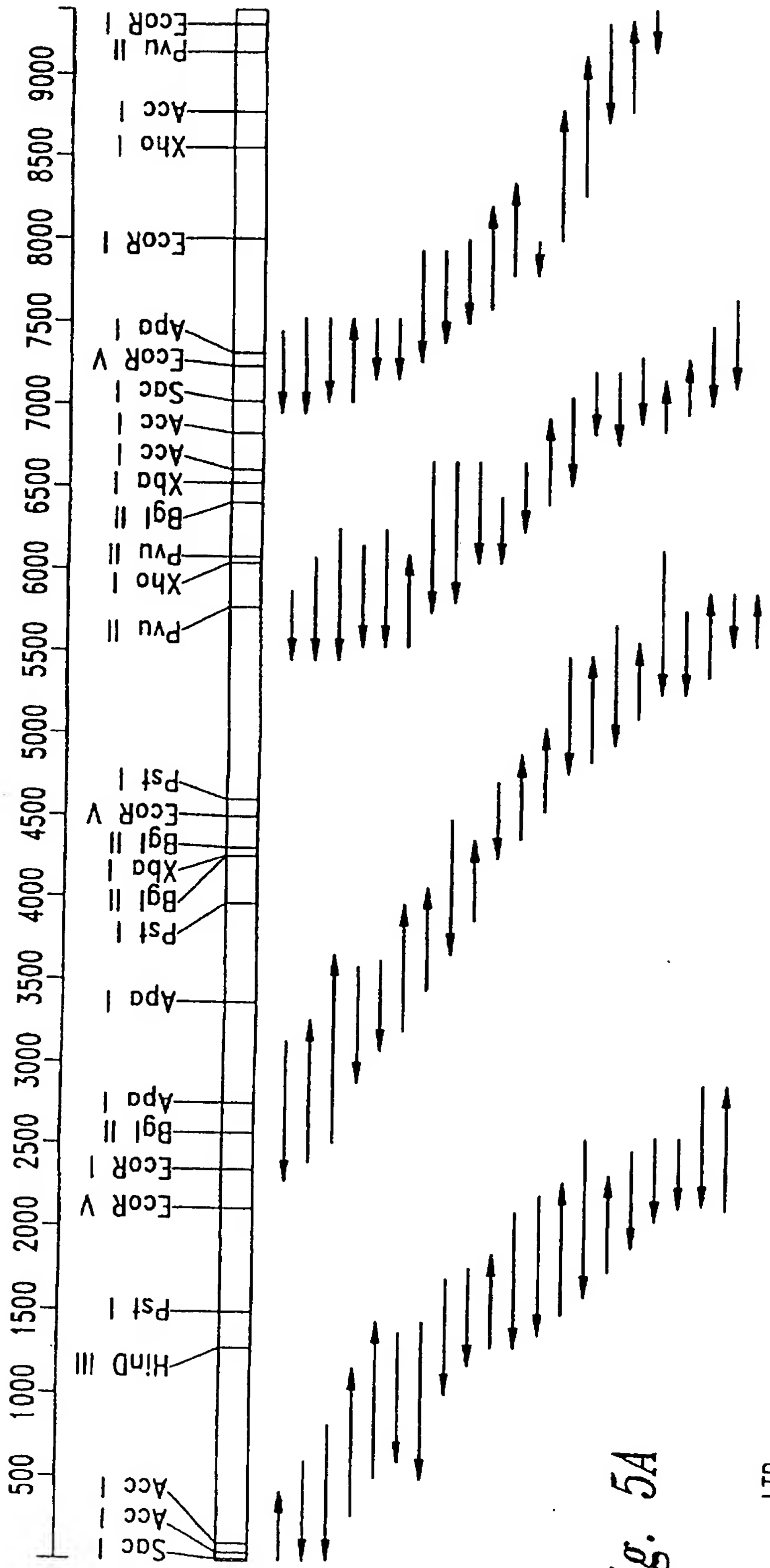

500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
Sac I
Acc I
Acc I
HinD III
Pst I
EcoR V
EcoR I
Bgl II
Apa I
Apa I
Pst I
Bgl II
Xba I
Bgl II
EcoR V
Pst I
Pvu II
Xho I
Pvu II
Bgl II
Xba I
Acc I
Acc I
Sac I
EcoR V
Apa I
EcoR I
Xho I
Acc I
Pvu II
EcoR I

Fig. 5A

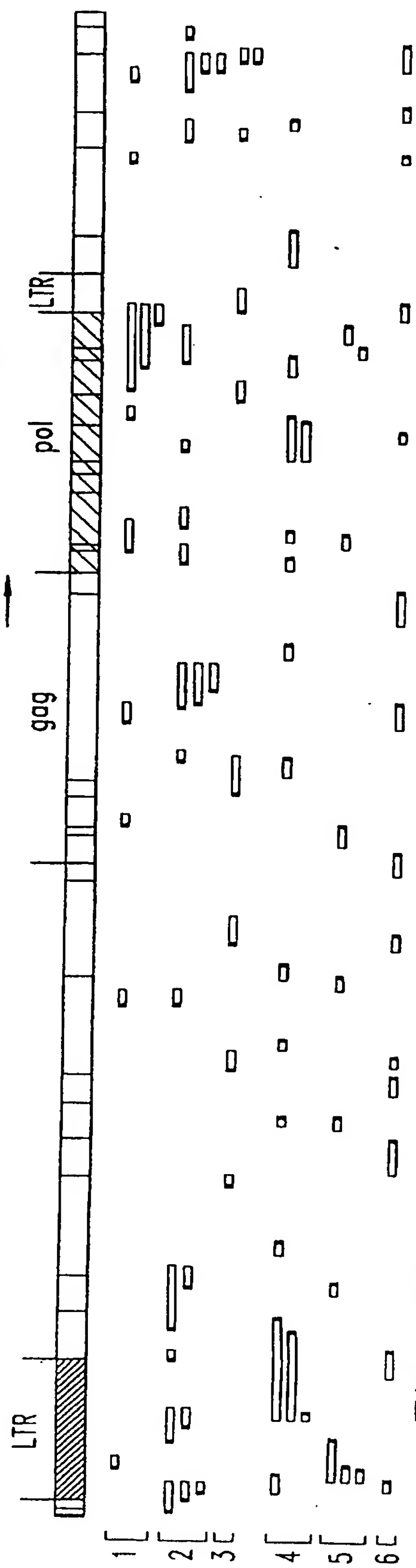

LTR
gag
pol
LTR
1
2
3
4
5
6

Fig. 5B

6/6

NUCLEOTIDE SEQUENCE OF THE REPRESENTATIVE
BREAST-TUMOR SPECIFIC cDNA B18Ag1

```
TTA GAG ACC CAA TTG GGA CCT AAT TGG GAC CCA AAT TTC TCA AGT GGA    48
Leu Glu Thr Gln Leu Gly Pro Asn Trp Asp Pro Asn Phe Ser Ser Gly
  1               5                  10                  15

GGG AGA ACT TTT GAC GAT TTC CAC CGG TAT CTC CTC GTG GGT ATT CAG    96
Gly Arg Thr Phe Asp Asp Phe His Arg Tyr Leu Leu Val Gly Ile Gln
             20                  25                  30

GGA GCT GCC CAG AAA CCT ATA AAC TTG TCT AAG GCG ATT GAA GTC GTC   144
Gly Ala Ala Gln Lys Pro Ile Asn Leu Ser Lys Ala Ile Glu Val Val
         35                  40                  45

CAG GGG CAT GAT GAG TCA CCA GGA GTG TTT TTA GAG CAC CTC CAG GAG   192
Gln Gly His Asp Glu Ser Pro Gly Val Phe Leu Glu His Leu Gln Glu
     50                  55                  60

GCT TAT CGG ATT TAC ACC CCT TTT GAC CTG GCA GCC CCC GAA AAT AGC   240
Ala Tyr Arg Ile Tyr Thr Pro Phe Asp Leu Ala Ala Pro Glu Asn Ser
 65                  70                  75                  80

CAT GCT CTT AAT TTG GCA TTT GTG GCT CAG GCA GCC CCA GAT AGT AAA   288
His Ala Leu Asn Leu Ala Phe Val Ala Gln Ala Ala Pro Asp Ser Lys
                 85                  90                  95

AGG AAA CTC CAA AAA CTA GAG GGA TTT TGC TGG AAT GAA TAC CAG TCA   336
Arg Lys Leu Gln Lys Leu Glu Gly Phe Cys Trp Asn Glu Tyr Gln Ser
            100                 105                 110

GCT TTT AGA GAT AGC CTA AAA GGT TTT                               363
Ala Phe Arg Asp Ser Leu Lys Gly Phe
        115                 120
```

Fig. 6

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/00398

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/48 C07K14/15 C07K16/10 G01N33/569 G01N33/574
G01N33/577 C12Q1/70 A61K39/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K G01N C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 88 01301 A (GEN HOSPITAL CORP) 25 February 1988 see page 4, line 4 - page 22, line 11; claims; figure 1 ---	1,3,4, 11-13, 21,22, 28,29
X	JOURNAL OF VIROLOGY, vol. 69, no. 1, January 1995, pages 414-421, XP002031129 SAUTER ET AL.: "Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas" see the whole document --- -/--	1,3-9, 15, 17-19, 24,26,27

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

3

Date of the actual completion of the international search	Date of mailing of the international search report
22 May 1997	30.05.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Macchia, G

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4 777 127 A (SUNI JUKKA ET AL) 11 October 1988 see the whole document ---	1,3-9, 15, 17-19, 24,26,27
X	BLOOD, vol. 86, no. 10 S1, 15 November 1995, XP000673682 WILLER ET AL.: "Transcription pattern of human endogenous retrovirus-K (HERV-K) related env sequences in patients with CML, NHL, colon cancer and mammal carcinoma" ab. 2236 see abstract ---	1,3-9, 11-13, 15, 17-19, 21,22, 24,26-29
X	WO 94 11514 A (ASTA MEDICA AG ;DIERICH MANFRED (AT); VOGETSEDER WERNER (AT)) 26 May 1994 see the whole document ---	1,3-9, 11-13, 15, 17-19, 21,22, 24,26-29
X	VIROLOGY, vol. 174, no. 1, January 1990, pages 225-238, XP000673688 WERNER ET AL.: "S71 is a phylogenetically distinct human endogenous retroviral element with structural and sequence homology to Simian Sarcoma Virus (SSV)"	2
A	see page 228 - page 230; figure 2 ---	10,14
A	GENOMICS, vol. 18, 1993, pages 261-269, XP000673683 LEIB-MÖSCH: "Genomic distribution and transcription of solitary HERV-K LTRs" see page 264 - page 265; figure 3 ---	10,14
A	VIROLOGY, vol. 209, 1995, pages 550-560, XP002031131 HALTMEIER ET AL.: "Identification of S71-related human endogenous retroviral sequences with full-length pol genes" see page 556; figure 6 ---	10,14

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCIENCE, vol. 234, 7 November 1986, pages 728-731, XP002031132 EARL ET AL.: "T-lymphocyte priming and protection against Friend Leukemia by vaccinia-retrovirus env gene recombinant" see the whole document -----	24,25

3

INTERNATIONAL SEARCH REPORT

.formation on patent family members

Interr nal Application No
PCT/US 97/00398

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 8801301 A	25-02-88	EP 0318502 A JP 2500323 T	07-06-89 08-02-90
US 4777127 A	11-10-88	NONE	
WO 9411514 A	26-05-94	GB 2273099 A	08-06-94